PRODUCTION OF EMPHASIS: THE PROGRESSIVE CONSTRUCTION OF PROSODIC CONFIGURATIONS IN CHILD DEVELOPMENT.

PACS Reference: 43.70 EP DEVELOPMENT OF SPEECH PRODUCTION

Marie-Ange Alexandre ; Claire Gérard
Laboratoire Langage et Cognition (LaCo UMR CNRS 6096)
MSHS
99, avenue du Recteur Pineau
86022 Poitiers Cedex
FRANCE
Tel. (33) 05 49 45 46 02
Fax. (33) 05 49 45 46 47
e-mail. Marie-Ange.Alexandre@mshs.univ-poitiers.fr

ABSTRACT
Conveying emphasis requires speech anticipation and speech control. Therefore, children may have difficulties in complex tasks such as reading texts aloud. Melodic, intensive and temporal speech variables must be structured so as to get specific configurations which enable the listener to focus his attention on target words. Acoustic measures were collected in experiments conducted on children aged 8 and 10 years old, as well as on adults. Data were processed through ANOVAs and a Correspondence Analysis. The two main axes resulting from the latter make it possible to clearly distinguish in the different age groups the progressive structuring and control of the prosodic variations conveying emphasis.

INTRODUCTION
Cognitive psychologists consider that mental representations are progressively built during the course of child development. This study attempts to observe distinct behavioural profiles or prosodic patterns which discriminate beginners from experts. The important steps we focus on are the inclusion of an effective management of the rhythm of speech, and the progressive increase of expressivity, as a function of the speakers' age and their speech styles. Of course, other processes must be taken into account in our paradigm, such as specific progress in reading and comprehension of read aloud texts, till they become fully automatic.

Researchers interested in the production of speech provided two main hypotheses: the first specifies the intervention of a prosody computer; the second supposes that there is a catalog of prosodic forms stored in long term memory, which are used as relative cues in different contexts by different speakers. We adopt the second approach.

Prosody has an expressive function (Fónagy, 1989), and allows the translating of the communicative intentions in global prosodic patterns, which can be decomposable in several prosodic groups. The concept of "prosodic group" was introduced by Vaissière (1980) to explain the fact that when two or several words are strongly linked by the meaning (i.e. an adjective and the word it qualifies), the prosodic contours of words are combined in a coherent way to form a single homogeneous acoustic pattern. The number of prosodic patterns of a sentence is dependent of the nature of the speech and of individual uses or speech styles. Morlec, Bailly & Aubergé (2001) consider the prosodic structure of a sentence as a superimposition of multi-

parametric prosodic contours, from the most global level (sentence, phrase) to the most local level (syllables, phonemes). The notion of prosodic pattern is fundamental for our study because the different acoustic parameters are interconnected. It was noted, for example, that an increase of fundamental frequency on a syllable is frequently associated with an increase of loudness. In the same way, an increase of loudness is often correlated with a lengthening of duration. The concept of prosodic configuration is also crucial for someone interested in the transmission of specific illocutionary forms (Gérard & Clément, 1998).

Emphasis has been chosen for two reasons. First, this illocutionary form seems to be decisive for speech comprehension, permitting to focus the listener's attention on the essential information within a message. Secondly, the emphatic stress appears to be more and more used by French speakers (Séguinot, 1976).

The Emphatic Stress

Two kinds of stress are frequently distinguished, as a function of their nature. The primary stress is a part of the morphologic level of interpretation: in french, it is composed of a duration lengthening and sometimes of a fundamental frequency increase (f0). It appears at a micro-melodic level. The secondary stress is associated with the enunciative, expressive and/or rhythmic function. It takes place in the macro-melodic variations. Some authors split up the secondary stress in three subclasses: the rhythmic stress (to escape stress collisions), the enunciative stress or focus stress or insistence stress, and the emphatic stress (Rossi, 1985). Other authors do not do this distinction and prefer to consider that the secondary stress, which is movable, can be placed on tool-words or on lexical-words (Pasdeloup, 1990).

In reference to Rossi's classification, we adopt the term "emphasis" but we envisage it in a less straight way, considering that processing emphasis is inevitably combined with other processes (lexical, semantic, syntactical) to form a body of emphases. Then it is necessary to approach emphasis in a multifactorial way. The acoustic characteristics of the emphatic stress, in French, are well known: reduction of the speech rate, insertion of silences before and/or after the target-word, increase of the pitch and/or the energy of the voice, introduction of melodic and/or loudness variations (i.e. Gérard & Dahan, 1995 ; Dahan & Bernard, 1994). By combining, these various modifications give place to more or less contrasted styles: some speakers use mainly the melodic variations, some others the temporal modifications, and a third group may mark emphasis by an association of different melodic, temporal and loudness variations. However, the temporal variations seem to be common to all the native French adult speakers. The introduction of silences before and\or after a target word (or group of words), as well as the reduction of the speech rate, are the most frequently observed phenomena (Grosjean, 1980 ; Gérard et Rigaut, 1994).

METHOD

Subjects

Three adults, four 8-year-old and six 10-year-old children were recorded. All the subjects were native French speakers, without any hearing deficit or regional accent.

Material

Four short texts (1 expositive, 3 narrative) were constructed, each developing two different sub-themes (connected to the same general theme): a brief introduction (from 2 to 4 sentences) presented the general story presented in the text. Each sub-theme was the object of one paragraph, and a short conclusion (from 2 to 4 sentences) referred to the general theme of the story. Each text was presented twice, with two different titles referring to each sub-themes of the text. The target-paragraph, which had to carry the emphasis, was framed. Target-words were pre-selected from a semantic-syntactic analysis in a pre-experiment.

The instructions underlined that it was necessary to speak naturally and to insist on the sub-theme selected, in order to be understood by a listener.

Procedure

After some minutes of mental reading, subjects read aloud the texts; and were recorded on a tape recorder. Each participant read each text twice, insisting alternately on each sub-themes. This procedure (semantic neutralization) made it possible to study prosodic variations within the same lexical and syntactic content. Adults read 8 texts (4X2) and children 6, the expositive text having been excluded for reasons of lexical difficulties.

Data Analysis

This analysis has two objectives: to show the existence of expressive strategies and to study the role of expertise on the relative weight of each kind of acoustic variations. The recordings were converted into spectrograms and segmented. Eight acoustic measures[1] were selected for every target-word and for both conditions of reading (with and without emphasis): the word duration (ms), the f0 mean, the f0 standard deviation and the f0 range (Hz), the loudness standard deviation and the loudness range (dB), a measure of the melodic contour of the word (Eady & Cooper, 1986) and a measure of the loudness contour (which refers to the same kind of information than the Eady and Cooper's measure).

ANOVAs (not presented here), attested of significant variations due to age and expertise for some of the acoustic measures, and a Correspondence Factorial Analysis (CFA) was done. The eight acoustic measures were introduced into the CFA. We operated a distinction within the values distributions, according to the direction of the modification operated by the speakers: increase or decrease of the value with regard to the condition of neutral reading. The correspondence analysis proceeded from the joint study of 16 lines-profiles (8 acoustics* 2 directions of the modification) and of 13 columns-profiles (the 13 participants)[2].

Results

Figures 1 and 2 present (1) the projection of the row-points in the system formed by the both axes revealed by the analysis (the main characteristics of these two axes are presented in appendix); (2) the projection of the column-points projection in the system of axes.

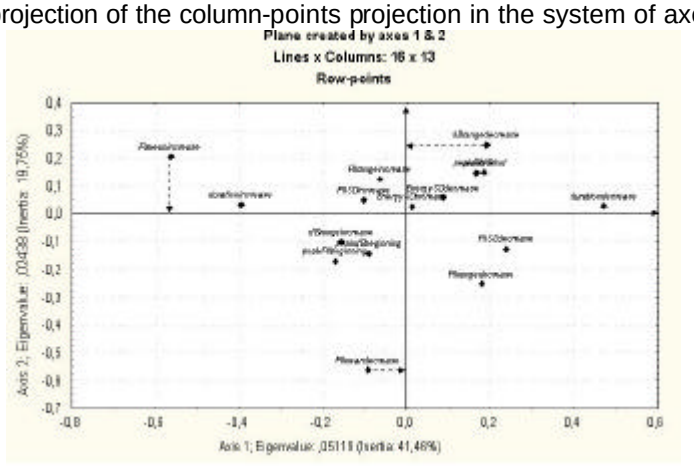

Figure 1: Projection of row-points.

[1] The mean loudness was measured but deleted because it distorted the cloud so much that it obscured everything.

[2] So, every quality (row factors) introduced into the analysis represented a specific acoustic variation (nature and direction). The raw data were transformed (homogeneity criterion). We counted the number of times the value measured in emphatic reading was superior (or inferior) to 10% of the value measured in neutral reading. Because the number of targets-words was not the same for the children and for the adults, we converted the data in frequency.

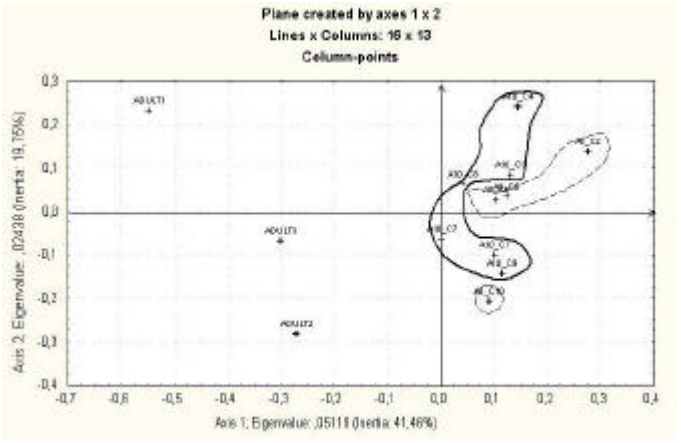

Figure 2: Projection of column-points (Individuals with their age and their number).

The first axis in figure 1 contrasts the typical qualities due to the absence or the presence of an emphatic stress:
On the right (the relative inertia are presented in brackets): decrease of words duration (21,4%), decrease of the f0 SD (7,8%), decrease of the loudness range (3,4%), peak of energy transferred to the end of the words (4,6%), decrease of the f0 range (4,1%), peak of f0 transferred to the end of the words (3,7%).
On the left: increase of f0 mean (24,8%), increase of words duration (18%), peak of f0 transferred to the beginning of the words (3,8%), increase of the loudness range (3,2%), increase of the f0 SD (1,8%). We interpreted the first axis as the illocutionary skill axis: it opposed the novices with the experts.

The second axis in figure 1 contrasts the use of the melodic modifications against the use of other modifications translating an effort in articulation:
On the top: decrease of the loudness range (12%), increase of the f0 mean (6,8%), peak of energy transferred to the end of words (6,2%), peak of f0 transferred to the end of words (6,2%), increase of the f0 range (5,7%).
On the bottom: decrease f the f0 mean (21,5%), decrease of the f0 range (17,4%), decrease of the f0 range (17%), peak of f0 transferred to the beginning of words (8%), peak of energy transferred to the beginning of words (6,4%). The second axis, less easy to interpret, seemed to oppose two main strategy (one based on melody and the other based on articulation).

DISCUSSION AND CONCLUSION

Reading aloud combines an activity of reception and treatment of a graphic (written) code with an activity of decoding and emission of an oral code. To read aloud a text generates a certain number of specific prosodic characteristics: as Vaissière (1980) noticed, loud reading is richer in the number of prosodic units than spontaneous speech, and inter-individual differences appear in the way of marking the prosodic structure. These differences produce strategies and different styles, referring preferentially either on fundamental frequency, or loudness and/or temporal modifications, as well as on a more or less high tendency to group the lexical words in the same prosodic shape. The prosodic structure of a read text seems very dependent from the spatial presentation of this text, at several levels. For example, pauses appear more systematically at the grammatical boundaries in the read text than in spontaneous speech. The visual marks structuring the text (such as paragraph, punctuation, underlining…) constitute constraints which condition not only the understanding of the text but also the rhythm of the reading. The reader must respect a certain number of interpretative imperatives ; his syntactic and semantic skills are essential in this situation. Then, the efficiency of the process of the text structure determines the kind of intonative management: global or local.

The first axis revealed by the CFA allows us to separate adults from children from the point of view of illocutionary skills. The axis 2 distinguishes two styles of speech. Several hypotheses can be advanced to explain this configuration of results. First, reading implies that the visual recognition of the words is fast and automatic, that the mental lexicon is developed enough, that the spelling, phonologic and semantic knowledge is pre-built. Learning to read forces children to develop strategies of deliberate memorising, in the verbal domain. Children become more and more able to manage the allocation of their cognitive resources in the required processes, and in the storage of the results: identification and memorisation correspond to a double task for children. When age improves, they have fewer difficulties with the task of identification, so that they can dedicate more resources to the storage of the information ; their performances improve. In this conception, it is justifiable to think that the task proposed to the participants is similar, for the children, to a double task. The children are still in the course of learning reading, and do not yet master the processes presented above, their reading strategies are not yet adaptable, and syntactic-semantic skills are less effective. These processes will demand progressively less control. An important part of the cognitive resources seems to be assigned too to the articulation effort produced with the aim of "well" reading, which reduces the part allocated to the understanding of the text (involved in the choice of the words which should carry the emphasis) and in the transmission of the emphasis. At the adult level, on the contrary, the processes involved in reading and understanding are perfectly automated ; so, a great part of the resources can be assigned to the convey of emphasis.

Two different expressive styles are observed in adults and children as well. For example, subject 1 (adult) is clearly distinguished from the other adults by its marginal use of the melody to convey emphasis. This subject has an experience of narrator for children which can explain this particular profile. As concerns infants, numerous studies have also described the use of a specific prosodic form adopted by adults when they talk to children (motherese). In our data, the two other adults mark the emphasis, on the contrary, mainly by transferring accents to the beginning of the words. In spite of these various styles, a major acoustic modification is used by the three adults and not used by any child: the slowing down of the speech rate on targets-words. For Fraisse (1967), in speech, time is the major structuring parameter which organizes mental representations in stable forms, adaptable to various contexts. We can consider that prosodic configurations get structured by means of the progressive control of the temporal processes, step by step, during the child development. The control of the speech rhythm is, for our conception, a key-stage in the progressive construction of prosodic configurations.

REFERENCES

Dahan, D. & Bernard, J.-M. (1994). Production et perception des phénomènes prosodiques dans la parole : comment identifie-t-on l'insistance malgré la variabilité inter-locuteurs ?, Journal de Physique IV, Colloque C5, Supplément au Journal de Physique III, Vol. 4, 501-504.

Eady, S.J., & Cooper, W.E. (1986). Speech intonation and focus location in matched statements and questions, Journal of the Acoustical Society of America, 80, 402-415.

Fónagy, I. (1989). On status and functions of intonation, Acta Linguistica, 39, Copenhague: Mienksgaard, 53-92.

Fraisse, P. (1967). Psychologie du temps, Paris: Presses Universitaires de France.

Gérard, C. & Clément, J. (1998). The Structure and Development of French Prosodic Representations, Language and Speech, 41(2), 117-142.

Gérard, C. & Dahan, D. (1995). Durational Variations in Speech and Didactic Accent During Reading, Speech Communication, 16, 293-311.

Gérard, C. & Rigaut, C. (1994). Patterns prosodiques et intentions des locuteurs : le rôle crucial des variables temporelles dans la parole, Journal de Physique IV, Colloque C5, Supplément au Journal de Physique III, Vol. 4, 505-508.

Grosjean, F. (1980). Linguistic Structures and Performance Structures : Studies in Pause Distribution. In H. W. Dechert & M. Raupach (Eds.), Temporal Variables in Speech, Studies in Honor of Frieda Goldman-Eisler, La Hague, Paris, NY: Mouton Publisher, 91-106.

Morlec, Y., Bailly, G. & Aubergé, V. (2001). Generating prosodic attitudes in French: Data, model and evaluation, Speech Communication, 33, 357-371.

Pasdeloup, V. (1990). Modèles de règles rythmiques du français appliqué à la synthèse de parole, Doctoral Thesis, Université de Provence, France.

Rossi, M. (1985). L'intonation et l'organisation de l'énoncé, Phonetica, 42, New York, Kargel, Basel: Kholer, 135-153.

Séguinot, A. (1976). L'accent d'insistance en français standard, In A. Séguinot (Ed.), L'accent d'insistance, Studia Phonetica, 12, Montréal, Paris, Bruxelles: Didier, 1-58.

Vaissière, J. (1980). La structuration acoustique de la phrase française, Annali Della Scuola Normale Superiore Di Pisa, Vol. X(2), T. Pacini-Mariotti (Ed.), 529-560.

APPENDIX

Table 1: The main characteristics of the two axes.

	Mass	Quality	Relative Inertia	Inertia Axis1	Cosinus² Axis 1	Inertia Axis 2	Cosinus² Axis 2
duration increase	,058661	,670305	,112509	,180671	,665854	,002535	,004450
duration decrease	,049180	,823735	,108217	,214224	,820832	,001591	,002903
f0 mean increase	,039897	,816813	,142878	,248736	,721861	,068697	,094952
f0 mean decrease	,016591	,434746	,100511	,002752	,011354	,215487	,423392
f0 SD increase	,089078	,445213	,020845	,018253	,363092	,008668	,082122
f0 SD decrease	,071302	,583008	,071668	,078456	,453921	,046846	,129087
Energy SD increase	,068141	,013272	,034092	,000233	,002833	,001802	,010439
Energy SD decrease	,054316	,164690	,029187	,007989	,113493	,007567	,051198
peak f0 end	,069524	,614583	,045243	,037578	,344400	,061898	,270183
peak f0 beginning	,067549	,583390	,054267	,038286	,292544	,079921	,290846
peak dB end	,069129	,425506	,074260	,046444	,259329	,062487	,166177
peak dB beginning	,076634	,433848	,040902	,012082	,122482	,064488	,311366
dB range increase	,066364	,791074	,023729	,031753	,554849	,028384	,236225
dB range decrease	,047403	,661899	,056950	,033815	,246204	,119877	,415695
f0 range increase	,090460	,730132	,019588	,007187	,152146	,057328	,577987
f0 range decrease	,065771	,787008	,065153	,041540	,264372	,172423	,522635