More than pointing with the prosodic focus: The Valence-Intensity-Domain (VID) model

Aubergé, V. & Rilliard, A.

Institut de la Communication Parlée UMR 5009 CNRS
INPG- Université Stendhal, Grenoble, France
{auberge, rilliard}@icp.inpg.fr

Abstract

This paper summarizes several perception experiments showing that the morphology of the prosodic focus conveys more information than the only deictic information: (1) the binary valence - yes/no focus – which is perceptively quite categorical (a magnet effect is clear on the basis of an identification and a discrimination experiment [2]), (2) the intensity information, used by the speaker to give his preference for one of two focused elements, (3) the information of the focus domain, that are some segmentation cues about the focused element (phonological unit or word unit), which are perceptively identified by listeners. The morphological cues revealing Valence-Intensity-Domain are observed in particular in morphing procedure making clear the thresholds of quite-categorical behaviors.

1. Introduction

Focalization can roughly be defined in the verbal stream as a function using prosody and/or syntax as a tool in order to bring out new information, or in order to contrast the enunciation. So does Rossi [9] consider focalization as a set of tools that allows speaker to hierarchize information and raise in the foreground one specific component by capturing the listeners' attention. The focus function is binary (there is or there isn't a focus - yes/no focus). The prosodic focus is a very robust binary function: a word focus is realized, at least in French, as a prominence on the first syllable (except in case of accents conflicts or style effect); however a word focus is perceived even outside this ecologic location of prominent syllable [4].The emphasis function by prosody is more often related to expressive cues as something like the degree of interest, and could be described in gradient terms. Jackendoff [6] showed the distinction between the "ordinary focus" versus the focalization on a syllable or a phoneme: the "metalinguistic focus". In previous study, we showed on the basis of an acoustic analysis that the F0, intensity and duration values of the meta-linguistic (syllable-focused item) focus vs. the contrast/new focus (word-focused item) are very similar, only the slope of the transition from the prominence to the low level is different [4]. We tried to show, by the way of perception experiments, that this three kinds of information are carried together by the morphology of the prosodic focus: the *Valence* (grammatical information about the focus: is it implemented or not by prosody – yes/no focus) carried by tonal processing, that is static cue, the *Intensity* (pragmatic information about the preference or emphasis) using a psycho-acoustic behavior (showed by Ladd [7]), and the *Domain* information (linguistic information about the focus function: contrast or new vs. intelligibility or phonological attention) carried by the slope of the contour, that is dynamic cue.

Thus, we will first briefly recall the results of a categorical perception experiment, held on stimuli progressively morphed from the no focus condition to the focus condition as produced by a speaker, and that makes clearly appear the binary decision of the yes/no focus information processed by the listeners into a magnet effect. In this experiment the extremes values of F0, intensity and duration give the basic level of the beginning of the gradient function of focus carrying the intensity information for which we then propose the preference function. Finally we present two experiments: the first one shows the perceptive competence of the listener to discriminate the domain (syllable vs. word); and from the second one, based on stimuli progressively morphed from the word focus condition to syllable focus condition, a threshold appears for some stimuli.

We can thus propose from these results which cues of the morphology of prosodic focus perceptively carry the valence, intensity and domain information, that is the VID model.

2. Valence evidences

The corpus used for the experiments presented in this paper (see [4] and [2] for more details) is based on a carrying syntactic structure, where lexical items vary only on the phonotactic dimension through length variations from 1 to 3 syllables. The French sentences recorded vary from 6 to 8-syllable length, and each lexical item varies from 1 to 3 syllables. The corpus was recorded with prosodic focus on each lexical item, and on each syllable for the meta-linguistic focus [4]. This corpus has been validated thanks to different perception tests [4] measuring the identification the focused item. The detailed protocol of these experiments is described in [2].

Four acoustic continuums starting from a neutral stimulus and ending in a focused utterance, with 8 intermediate steps, were constructed by means of analysis-resynthesis technology. The frequency, intensity and durations steps used in this morphing procedure are all smaller than the perception threshold described by Rossi [8]. This identification task was intended to reveal a quite categorical behavior of prosody: a magnet effect. Results are comparable to those given by Ladd [7]. Results are presented in the figure 1.

Answers were tested thanks to a Probit analysis: the slope of the identification curve and the value of the threshold over which the answer falls over in the other category (focus or non-focus) are summarized in the figure 2.

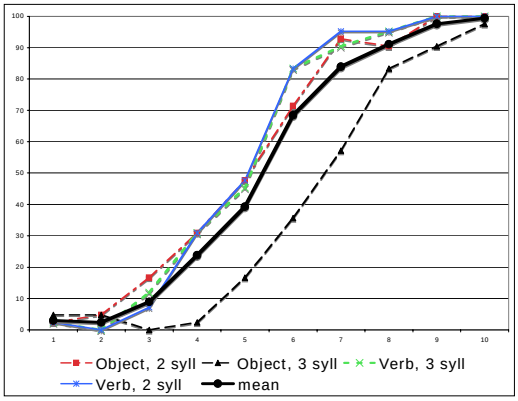

Figure 1: results of the identification percentage of focus in the four continuums, plus the mean percentage. Ordinates correspond to the 10 iterations of the morphing from the neutral stimulus (1) to the focalized one (10).

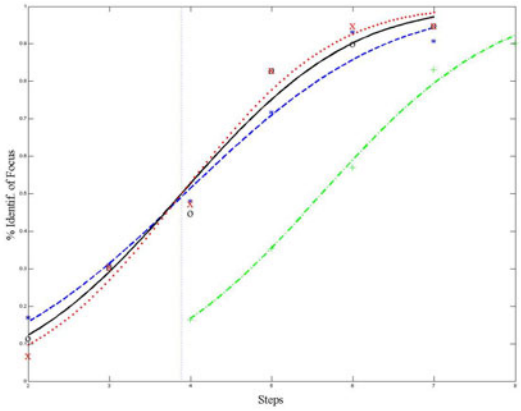

Figure 2: plot of the probit analysis. The 4 curves are the results of an interpolation of the identification percentages for each sentence. Legend: stars (*) the dashed line represent 2-syllable object stimuli, (+)the dash-dotted line represent the 3-syllable object stimuli, (o) the plain line represent the 3-syllable verb stimuli and (x) the dotted line the 2-syllable verb stimuli. The vertical line around the 4th iteration represents the threshold for the three grouped curves. Coordinates represent the identification score, and ordinates the iteration number.

To complete this identification task in order to show a categorical behavior, we held then a discrimination experiment, but we could not observe any maximum of discrimination. We should then conclude to a magnet effect of yes/no focus, with a clear static F0 level (see figure 3) for this speaker (intensity behave in the same way).

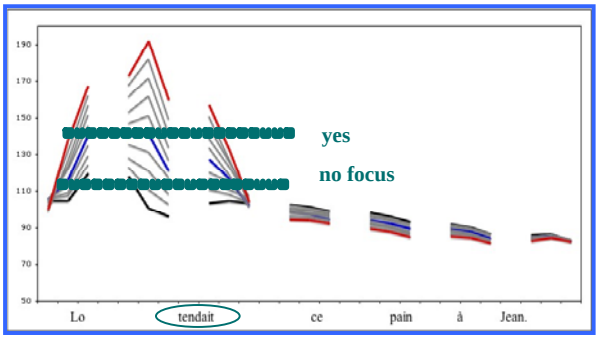

Figure 3: the F0 morphology of the focus boundary valence for the 2-syllable verb. The lowest and highest levels came from the two natural references pronounced by the speaker.

3. Intensity evidences

Ladd [7] showed that in a psycho-acoustic condition, the listeners perceive the gradience for varying prominence (only on F0 varying stimuli), in opposition with a linguistic task, in which he could show a categorical perception behavior, almost the same that we noted in the preceding experiment. In preliminary experiments, we could observe that the speakers and the listeners use this psycho-acoustic ability to give or recognize their preference between two focused items that is presented in the same utterance. Such distinction is typically used in human-machine dialog, when a choice is proposed to the listener, by adding the listener's preference about this choice. We gave to listeners utterances such as: "Do you want to travel through *Paris* or through *London* next week?", with "*Paris*" and "*London*" raised over the perceptive threshold of focus. According to the previous experiment both words are supposed to be focalized, but one of these two pointed items is more "intense": this one is recognized by the listeners as the choice proposed by the speaker. This experiment needs to be reproduced with a varying level of intensity and by controlling the place of presentation, since it can been expected that the first presented stimulus could be chosen as the preferred one, like an orthogonal cue to the intensity of prosody.

4. Domain evidences

4.1. Discrimination experiment

The listeners have to judge whether the speaker is talking more specifically of one person, one action or one object (in contrast with another one), or if he was misunderstood and thus repeat the bad understood syllable (meta-linguistic focus). 25 listeners listened only once each stimulus and answered on the screen presented in figure 5. The stimuli were selected for all the length and places of word focus, and their corresponding stimuli for the syllable focus on the first syllable of the word, since in French (like in many languages) the word focus is realized in general with a strong prominence on the first syllable and must be compared with the stimuli produced with a focus on the first syllable.

Figure 5: test interface of the perception test for syllable vs. word focused stimuli

It must be noted that the listeners said that they had the feeling to answer by hazard. But the results, presented in table 1, show that the listeners can discriminate significantly largely over chance if the speakers gave a focus information about the whole word (contrastive focus) or only on the first syllable (meta-linguistic focus), that corresponds to two completely different communicative functions.

Table 1: Percentage of right answers for the word vs. syllable identification task

	% good answers	% good answers corrected from the chance level
word focus	79,7	59,3
1st syllable	83,7	67
Total	81,7	63,3

It must be noted that the acoustic analysis (see [4]) performed on these stimuli showed that the F0 and intensity levels are the same on the first syllable for both the word focus and the first syllable focus task. The values of values and intensity inside the first syllable of the 1st-syllable-focused items and the word-focused items are similar. It was indirectly confirmed in our experiment on the focus level identification; since the same threshold was found whatever the speaker was performing a word focus or a syllable focus. It means a priori that the cues of this domain identification must be found in dynamicity of the contours between the first (where in prominence, both in word and first syllable focused stimuli) and the second syllable of the stimuli (see figure 5). The following experiment is held in order to know if the identification of the domain is categorical or continuous, and in order to determine where is the morphological boundary in case of quite categorical perception.

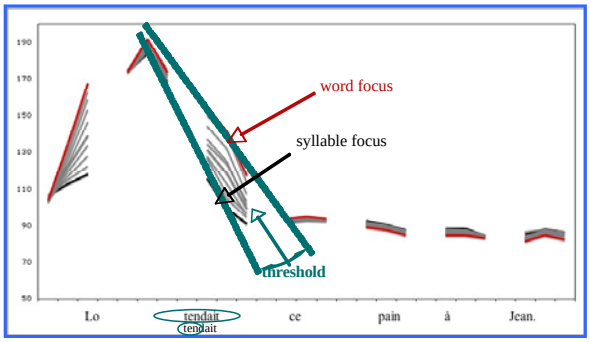

Figure 5: the F0 morphology of the focus boundary domain for the 2-syllables verb. The less and more abrupt contours are the two natural references pronounced by the speaker

4.2. Identification experiment

4.2.1. Stimuli

The stimuli used for this experiment rely on six different sentences, all based on the same Subject-Verb-Object syntactic structure. For each sentence, one word only (either the subject, the verb or the object) is focused. The length of the focused word varies from 2- to 3-syllable length; all the other words are monosyllabic ones. A native French speaker has recorded these sentences. He was asked to perform the sentence first with a contrastive focus on the given word and then with a meta-linguistic focus, only on the first syllable of the same word.

As for the focus identification experiment, the stimuli of this domain identification experiment have been constructed thanks to the Praat software [3]. The prosodic parameters (i.e. fundamental frequency, duration and intensity) were gradually scaled, using 8 intermediate steps from the word-focused stimulus to the 1st-syllable-focused stimulus. It results in 10 stimuli for each sentence, resulting in 60 different stimuli. The frequency, intensity and durations steps used in this morphing procedure are all under the dynamic perception of glissando, as described by Rossi [8].

4.2.2. Experimental protocol

Listeners heard all the stimuli in a different random order for each subject. Each stimulus is proposed three times during one listening session. They have to answer if they think that the speaker intended to contrast one word (the subject, the verb o the object) or to be very intelligible on the first syllable of one of this word.

11 listeners, all native speakers of French without any hearing problem, participated in this test. They can listen to the stimulus only one time and have to give they answer (word focus or syllable focus) thanks to computer interface.

Table 2: Results of the ANOVA analysis of the word to syllable focus perception test. The factors are the 10 steps of the continuum (Iteration), the Length of the focused word, the position of the focused word in the sentence and the 3 rep.

Factor	ddl	F	p	sig.
Iteration	9	64,129	,000	*
Length	1	4,790	,053	
Position	2	9,981	,001	*
Repetition	2	1,067	,363	

Iteration * Length	9	14,345	,000	*
Iteration * Position	18	7,168	,000	*
long * Position	2	1,785	,194	
Iteration * Length *	18	3,599	,000	*

4.2.3. Results analysis

First the consistency of the listeners answer was check: the Cronbach's alpha on the result is of 0.84.

An ANOVA analysis was held on these results, in order to check the relative influence of the different factor involved in the experiment: the iteration on the word-focus to 1st-syllable-focus continuum (10 steps), the length of the focused word (2 or 3 syllables), the position of the focused word in the sentence – linked to its syntactic function (Subject, Verb, Object) and the three repetitions of each stimulus. The major results of this analysis are summarized in the table 2 and in the figure 6 to 9.

The factors that can explain the major part of the variance in the results are (1) the continuum from the word-focused stimulus to the 1st-syllable-focused one, (2) the position of the the word in the sentence – either Subject, Verb or Object position, and (3) the interactions between the 10 steps and the length, and the positions. The 3 repetitions do not affect the listeners' answers. Such results are a confirmation of the ability of listener to perceive the focused domain, and then to distinguish between the two underlying functions: contrastive focus or meta-linguistic focus.

The strong effect of the position of the focused word on the listeners' answers raise some question about the influence of some other factors that were not taken into account in this experiment: the phonological structure of the focused syllables, and the nature of its component (e.g. voiced / unvoiced consonant) and the relative influence of the position and of the nature of the focused word on the perception of focus.

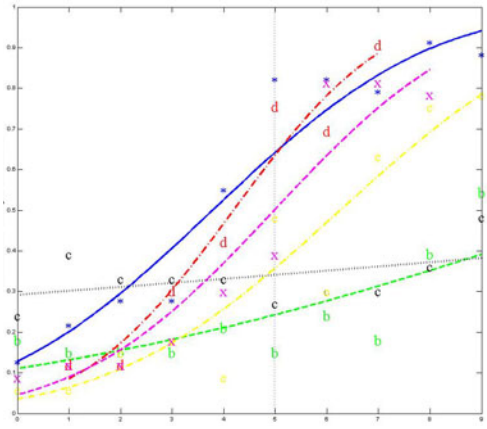

Figure 7: results of the Probit analysis. The 6 curves are the results of an interpolation of the 1st-syllable-focus identification percentages for each sentence. Legend: (*) is 2-syl object stimuli, (b) dashed line is 2-syl subject, (c) doted line is 2-syl verb, (d) dash-dotted is 3-syl object, (e) dasd-dotted line is 3-syl subject (x) dashed magenta line is 3-syl verb. The vertical line around the 5th iteration is what could be the threshold for the curves except b and c.

The effect of the stimuli's length in interaction with the iteration show that the longer the word is, easier the distinction between word and 1st-syllable focus is. But, as this effect is not significant alone, and as 2-syllable length object word received good results, it can be question if it is the consequences of the stimulus' size or of other factors (e.g. those already listed above).In order to check if the pattern of answer is categorical or continuous, a Probit analysis was performed on this results. Results are summarized in the figure 7. This analysis shows that 4 out of 6 stimuli show an abrupt increase of 1st-syllable-focus answer, around the 5th iteration. For the two other stimuli either the syllabic focus is not recognized, or only at the very last step. These results are coherent with the preceding ANOVA analysis, as some stimuli don't receive good identification scores, and the reasons of such a behavior have to be investigated. But the important result of this analysis is that for almost some stimuli, listener did answer as if there was a boundary between the word and the syllabic focus.

5. Conclusions

We held several experiments in order to build the evidences about the complex information carried by the prosodic morphology of "focus": Valence-Intensity-Domain. We pointed on the quite-categorical processing, and the threshold between pseudo-categories, of valence (which linguistic functions of contrast or new are well studied) and domain (word vs. syllable, which implies another – meta-linguistic – function), and how the gradience perception behavior (shown for example by Ladd [7]) is, for example, used for the speaker function of preference. We proposed that the valence uses some static cues (with an identical processing whatever the domain), the intensity uses the gradience perception inside the focus pseudo-category, and the domain uses the dynamic of the contours between the first and second syllable (the threshold could be the glissando psycho-acoustic ability), in order to give the cues of the contour being global to the word vs. global to the syllable. We are now under implementing this model in the France Telecom R&D TTS/dialog system in order to evaluate the relevance the static and dynamic thresholds deduced from the perceptive experiments.

6. Acknowledgment

Many thanks to Philippe Bretier of France Telecom R&D, for fruitful discussions about the preference function in dialogs.

7. References

[1] Amir, N., Almogi, B., Gal, R. 2004. Perceiving Prominence and Emotion in Speech – a Cross Lingual Study, *Speech Prosody 2004*, Nara, Japan, 375-378.

[2]Aubergé, V. 2001. Modélisation de la prosodie par formes globales : amont ou aval de la phonologie tonale ?. *23rd JEP* , France, 281-284.

[3] Boersma, P. & Weenink, D. 2005. Praat (Ve4.3.01) Ret Feb 9, 2005, from http://www.praat.org/

[4] Brichet, C., Aubergé V., 2002. La prosodie de la focalisation en français : faits perceptifs. 94-99, 24es JEP, Nancy.

[5] Fonagy, I. 1983. *La vive voix, Essais de psycho-phonétique*,, Payot.

[6] Jackendoff, R. 2002. *Foundations of language*, Oxford: Oxford University Press.

[7] Ladd., D.R. & Morton, R. 1997. The perception of intonational emphasis : continuous or categorical ? Journal of Phonetics, 25, 313-342.

[8] Rossi, M. 1978. La perception des glissandos descendants dans les contours prosodiques. *Phonetica*, 35, 11-40.

[9] Rossi, M. 1985. L'intonation et l'organisation de l'énoncé, *Phonetica*, 42, 135-153.