

HOW SPEAKERS AGREE IN ENGLISH: AN EXPERIMENTAL TOOL FOR ANALYSING PROSODY AND GESTURE SIMULTANEOUSLY IN SHORT FILM SEQUENCES OF CONVERSATIONS

Laurence Delrue

STL, UMR 8163 - Université de Lille 3, France

laurence.delrue@univ-lille3.fr

ABSTRACT

Our aim was to develop an experimental methodology, leading to a joint approach to verbal and non-verbal activities in discourse. Authorware provided the suitable authoring environment to design a semi-automatic application, making possible the acquisition of a numerical database of prosodic and kinetic variations in short filmed sequences of dialogue, converted into graphic representations. This method of pluriparametric analysis brings out the interaction between specific intonation patterns and associated explicit kinetic markers. It enlarges our focus of interest from the isolated sentence structure to a wider field of study, involving gestural behaviour associated to prosody, going beyond the word to affect the entire speech act. This study will focus on the complex combination at work in the dynamics of interaction.

Keywords: experimental technique, intonation, gesture, discourse, interaction

1. INTRODUCTION

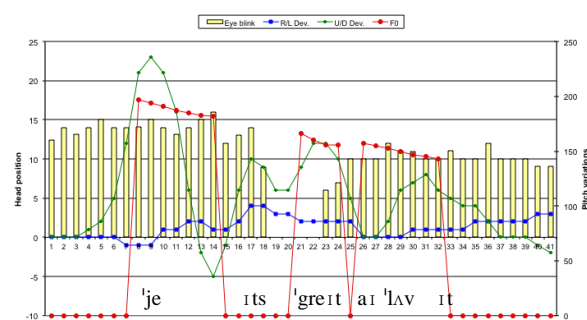
Analysing prosodic features and gestures in a synchronous manner is not an easy task. It seemed therefore interesting to create an experimental tool to submit both prosody and gesture to a simultaneous analysis. Our aim was to give a reliable account of the organization of spontaneous speech in face-to-face interaction in an interview setting.

2. METHODOLOGY: FROM DATA COLLECTION TO GRAPHIC REPRESENTATION

Conversations between young native English speakers were recorded and filmed in a studio by two cameras placed at right angles (one filming the speaker's face and the other his/her profile). The speakers knew each other well, thus ensuring that the speech data were as natural as possible in a

filmed environment. The chosen extracts were labeled using a time-frame code and concatenated into short Quicktime sequences easily accessible on a computer screen. The Fundamental frequency, (F0) analysed with Praat (speech analysis software), was then synchronized to the film format in order to have the numerical pitch values appear on a final graph together with the kinetic data. The acquisition of such data is obtained from the face sequences: each speaker agreed to have a small black dot placed between their eyebrows which could be clicked upon image by image; each head deviation was thus recorded and the numerical values of both right/left and up/down movements entered in an Excel chart and then shown on the final graph. The values of the head movements were thus semi-automatically measured from the first to the last image of each sequence and are represented analogically on both a horizontal and a vertical axis. The profile sequences were also analysed to measure opening and closing movements of the eyes during interaction. All data were finally concatenated from an Excel chart by Authorware to provide graphic representations such as the following:

Figure 1: Graphic representation of an utterance showing F0 variations in Herz; degree of eye opening appearing as histograms, up/down head deviations as rhombi and right/left head deviations as squares.



3. PLURIPARAMETRIC ANALYSIS: TOWARD A TYPOLOGY OF ASSENT

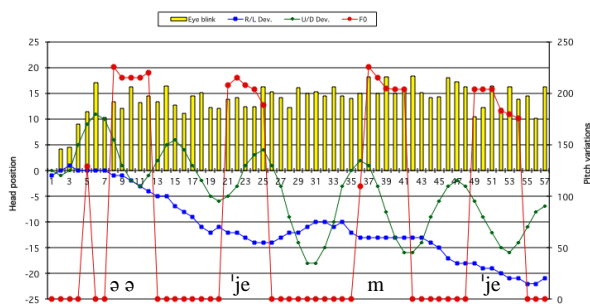
3.1. Assent with *yes*

Phonetic and kinetic data thus synchronised and given a visual analogical representation give a pluriparametric insight into the speech act itself. This methodological lead was carried out on a group of answers to questions asked during the interviews. Figure 1 above shows a speaker's answer "Yes, it's great, I love it." to the question asked by speaker A: "Do you like it here in France?". Speaker B thus gives an affirmative answer, using three syntactically assertive units, realized phonologically as three tone units made up of three falls of gradually declining range (measured in Herz). The graph shows how the head-nods (measured in pixels) are not only perfectly synchronous rhythmically to each pitch variation but also vary in scope in the exact same way as the intonation pattern to form a decrescendo. Studying the combined effects of verbal and non-verbal cues in meaning production thus leads to extending the notion of nucleus from its intonative framework to an all-inclusive concept integrating kinetic deviation.

3.2. Positive feedback with *yes*

Similar verbal forms, of the *yes* type, may also be used by a speaker not as an assent as such but to give positive feedback to a co-speaker, as in "uh, uh, yeah, hum, yeah", where the speaker alternates level tones on gap fillers with low falling tones with an assertive value [2] on both *yeah* items, as shown in the following graph.

Figure 2: Positive feedback: coordination of phonetic and kinetic markers to show agreement with the other speaker during interaction.



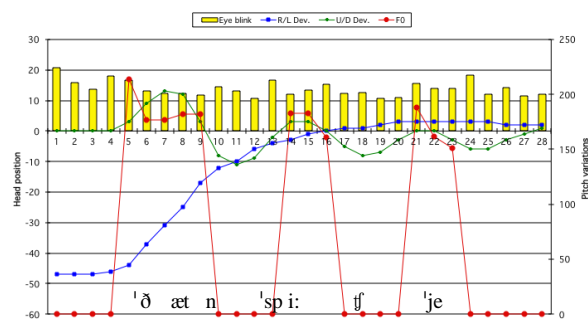
Each verbal form of the utterance is synchronized with a vertical nod of the head; the scope of the nods is minor on the gap-fillers but they are followed by a major up/down deviation on

central "yeah", giving the word a distinctly marked status. It is followed by two additional kinetic and phonetic nuclei, prolonging the positive feedback effect.

3.3. Confirmation with [(X) and (Y)] = *yes*

A third type of assent can come in the form of an answer to an utterance such as: "So you were studying drama?" In spite of its syntactically assertive organization, the unit is assigned a rising tone and is thus transformed into a pragmatic query, the speaker expecting confirmation. Speaker B thus answers "That and speech, yeah."

Figure 3: Giving confirmation: "that and speech yeah."



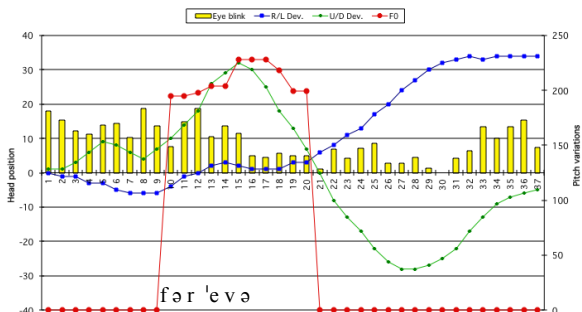
As shown on the graph, speaker B takes up the information just given by speaker A, using an anaphoric deictic (reference to *drama*), coordinated by *and* to a new item of information (*speech*), both lexical items becoming direct objects of the elliptical verb *study*. Each item receives a low falling tone giving each of them an individual status as a separate piece of information carrying an assertive value. To each falling tone is synchronized a vertical nod of equal dimension (13/12 and 14/11 pixels). Here again synchronicity and scope are identical on both levels, phonetic and kinetic; the final item *yeah* is assigned both a falling pitch variation and a nod, but of minor scope on this occasion. It comes as a final positive confirmation in the linearity of the sentence organization, which applies to both the old and the new item in the same assent-giving speech act, both items having already received phonetic and kinetic confirmation as separate units.

3.4. Repeat: assent without *yes*

However, some speech acts expressing assent may simply do away with a linguistic form exclusively devoted to agreement per se. A speaker may pick up another speaker's last word in interaction, as is

the case here with the word *forever* appearing as an isolated unit on the following graph:

Figure 4: Repeat and agreement on co-speaker's final item: "forever."



This single lexical item is an assenting speech act in itself: the speaker assigns the word a low falling tone, which is perfectly synchronous to a characteristically marked nod (28/60 pixel up/down deviation). Pitch variation and gesture are combined to give the word the meaning of an assent in interaction, although the verbal form *yeah* is now lacking. The speaker agrees with her co-speaker on his last term and ends the topic of conversation at this point. The finality of the speech act is visible on the graph since, when the speaker stops speaking, she changes positions and moves to a listening position - no up/down or right/left deviations being the kinetic cues to turn-taking.

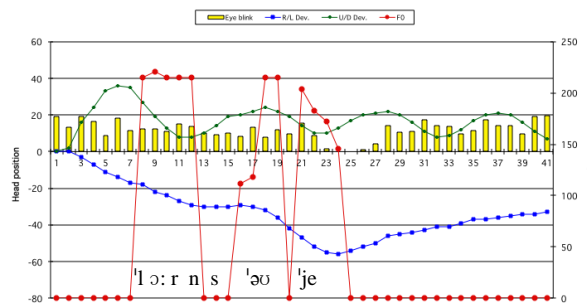
3.5. Repeat question with yes

A variation of the same type of assent comes when a speaker repeats what her co-speaker has just said but under the form of a question which she answers herself, saying "Lawrence? Oh, yeah." The first item, the name, bears a mid-level tone giving it a continuative value and *yeah* bears a falling tone. The general kinetic set up starts with a marked vertical nod on *Lawrence*, showing that the speaker's intent here is, in fact, not to repeat in order to ask for confirmation but to select one item out of many, and to agree with the co-speaker. This major assenting kineme is followed by a general gestalt made up of three minor nods.

The speaker thus shows her total agreement to what her interlocutor has just said. The broader interactive context shows that they share the same knowledge concerning the third person they are talking about (namely, his height, which is not a matter for discussion). The verbal form *yeah* and its accompanying head-nods come as an echo to

the first emphatic and contrastive kinetic and phonetic nucleus. They are not redundant as such, but simply part of the agreeing process in oral communication (repetitions, false starts, hesitations, etc.).

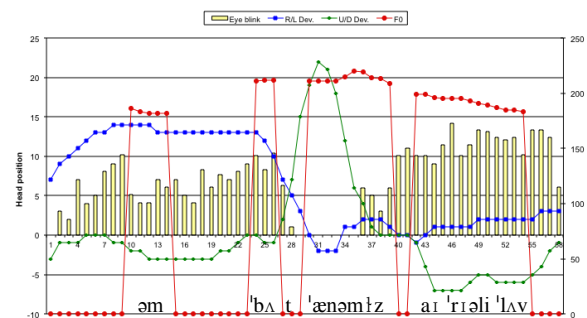
Figure 5: Argument: "Lawrence? oh yeah."



3.6. Assent as the backbone of the argumentative structure

A speaker, talking about her personal likes and dislikes, says: "You know, I'm not around children, I don't look after children, I get impatient with children, uhm, but animals I really love."

Figure 6: Argument: "uhm, but animals I really love."

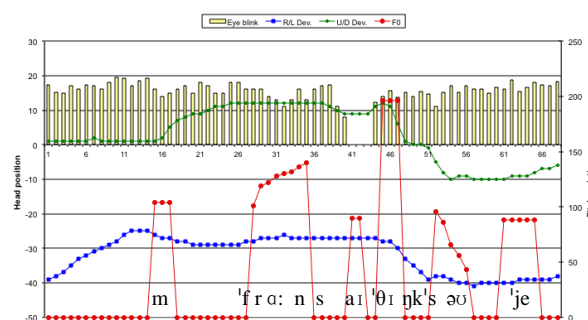


The two nuclei, *around* and *after*, are assigned continuative mid-level tones and horizontal head-nods indicating negation. The word *impatient* bears a fall (ending the list of negative items) and a vertical nod, punctuating and giving an emphatic value to the semantically negative notion. In the final Intonation and Gesture unit, after two minor vertical nods on *uhm* and *but*, the speaker assigns a contrastive focus to the word *animals* (as opposed to *children*) by disrupting word order structure but also by using marked forms both in intonation and gesture. The graph shows the synchrony of all forms in a major assent hub (falling tone, peak in intensity, major vertical head-nod) followed by nods gradually decreasing in amplitude and associated with a final declining intonation curve.

3.7. Assent but with a restriction: where kinetic cues anticipate a final ‘I don’t know’

As answers to the questions asked in the interviews get more complex, the prosodic and kinetic cues are also more intricately associated in a complex network of meaning. As is the case in the following answer to the interviewer’s question “Would you like to live here?”, where the English speaker says: “In France? I think so yeah, I haven’t decided yet, I’m still quite young, so I don’t really know”. The first intonation unit (*In France?*) is assigned a rising tone synchronous to a rising head movement, typical of questions, by which the speaker expresses his need for confirmation of the place referred to before answering. The second IU (*I think so*) is pronounced with an expected high fall on *think*, accompanied by a marked nod (up to down movement of 20 pixels) continued on post-nuclear *so*, clearly marking a determined assent. However, *yeah*, which should function as a reinforcing echo of the former assent, and should be an assent itself, is here given a mid-level tone which is not synchronous with an expected affirmative head-nod: the speaker seems to stand completely still at that point. No deviations are visible on the graph as well. Moreover, he noticeably squints as he utters the word *think*, finding it visibly difficult for him to answer the question in too straightforward a manner. The kinetic cues and the level tone on the assent word *yeah* indicate that there is more to come, and that the first assent may need some nuance. The next tone units both bear assertive low falling tones (*I haven’t decided yet, I’m still quite young*); their function is to explain the difficulty expressed kinetically on *think*. On the word *young*, the speaker starts moving his head from left to right, thus anticipating the negative clause to come (*so I don’t really know*) bearing a final terminative fall and being punctuated by four small repeated left-right negation movements. The speaker does not actually associate the fact of being young with a negative head movement, which may have been the case if the IU were taken in isolation, but indicates by the kinetic overlap that the cause of his restricted assent is his age. The head movements associate the word *young* and the conclusive statement in a relation of consequence, expressed verbally by the word *so*.

Figure 7: Assent with a restriction: “In France? I think so yeah, I haven’t decided yet, I’m still quite young, so I don’t really know”



4. CONCLUSION

By making it possible to study the combined effects of verbal and non-verbal parameters, the methodology presented here enlarges the focus of interest from the isolated word or even sentence level to a wider field of study, involving gestural behaviour associated to prosody going beyond the word to affect the entire speech act. This study is but a small contribution to the study of how prosody and gesture produce meaning together in spontaneous speech. To quote A. Kendon [1]: “As a close examination of the coordination of gesture with speech suggests, these two forms of expression are integrated, produced together under the guidance of a single aim.”

5. REFERENCES

- [1] Kendon, A. 2004-2008. *Gesture: Visible Action as Utterance*. Cambridge University Press.
- [2] Wells, J.C. 2006. *English Intonation, An Introduction*. Cambridge University Press.