RELATIONSHIP BETWEEN GESTURES AND VOICE IN VERBAL INTERACTION: PROSODIC AND KINESIC ASPECTS OF BACK-CHANNEL SIGNALS

Roxane Bertrand, Jacques Boyer*, Christian Cavé*, Isabelle Guaïtella* and Serge Santi***

**Laboratoire Parole et Langage, U.R.A. 261 CNRS*
Université de Provence, Aix-en-Provence, France
*** Laboratoire de Phonétique, GRI-DESYCOLE,*
Université de Franche-Comté, Besançon, France

ABSTRACT

This paper points out the theoretical merits of (1) investigating the relationship between vocal and gestural activity in speech, and (2) using this perspective to study back-channel signals (which can be verbal, vocal and/or gestural). The main results of two preliminary studies on back-channel signals are reported.

1. THEORETICAL BACKGROUND

1.1 Three-modality communication and the relationship between gestures and voice

In a trimodal model of communication [1, 20, 11, 14], interpersonal exchanges are based on three communication modalities: verbal, vocal, and gestural. It was hypothesized here that among these three modalities, vocal and gestural activities are tightly linked [14, 15]. Some phoneticians contend that the gestures which co-occur with speech are linked to intonation by their temporal features, and above all, by their semiotic characteristics [17, 7, 14, 15]. The study of the relationship between eyebrow movements and variations in fundamental frequency has shown that, although the interaction between the two depends on the context and the speaker, these two kinds of movement are clearly synchronized [9, 16, 2]. Phoneticians working on phonatory gestures and their interaction with expressive phenomena [5, 6, 12, 13, 21, 22] have contributed to our understanding of the link between vocal and gestural expression in speech. If we agree that the expressivity of the voice is related to a "glottal movement", i.e. that the voice is the audible trace of a physiological activity, and that such "internal gestures" are comparable to "external gestures" such as facial expressions, then we can assume that there is a link between the visual and auditory channels. The lack of a link would be surprising in that it would reflect the disconnection of internal and external gestures. Some researchers in non-verbal communication or human ethology [4, 10, 11] have addressed the dual question of the micro-analysis of gestures (in Condon's terminology) and their link to the vocal component. The problem of bimodal perception can also be considered relevant to this line of research.

1.2 Conversational Feedback

During conversation, the listener contributes to the interaction by exhibiting an active listening attitude, or by showing his/her desire to speak through the production of specific signals [18, 23, 19]. It is well known that turn-taking is controlled by listener-produced feedback called back-channel signals. Back-channel signals can be vocal, verbal, gestural, or any combination of the three. The acoustic characteristics of vocal back-channel signals have not been sufficiently described (see however [24]). Likewise, little research has been conducted on the forms and functions of gestural signals [8]. Such studies could help us gain insight into the relationships between the forms and functions of back-channel signals, and allow us to suggest a precise and objective typology.

2. EXPERIMENTAL STUDIES

Two studies were conducted to describe the formal and functional characteristics of vocal and gestural back-channels signals. In the first study, the prosodic characteristics of vocal signals were analyzed in relation to their functions. In the second study, the same types of analyses were performed, but gestural activity was also taken into account.

2.1 Experiment 1: Prosodic and Functional Analysis

In this preliminary study [3], the prosodic and pragmatic aspects of back-channel signals in turn-taking were investigated. Most of the observed signals (10 out of 15) had a flat or slightly falling prosodic contour. The listener appears to use these signals to show that he/she is listening but does not wish to interrupt. Among the other five cases were two repetitions of the speaker's utterance. These signals had rising contours and can be interpreted as questions. There were also two isolated signals whose function appears to be to prompt the speaker to continue. For the remaining case, we do not have a functional interpretation to propose.

As a whole, the prosodic contours of the listener's signals were inverted with respect to the contour of the preceding speaking turn. In other words, back-channel signals with rising prosodic contours follow utterances with an overall falling intonation, while those produced with flat or falling contours follow utterances whose overall pattern is rising. Thus, two types of vocal back-channel signals can be distinguished: (1) those with a rising prosodic contour, which appear to have a continuation function, and (2) those with a flat or falling contour, which manifest an active listening attitude.

2.2 Experiment 2: Prosodic and Gestural Analysis

In this study, the prosodic analysis of vocal back-channel signals was extended by an analysis of the gestural feedback produced by the listener (Boyer, doctoral dissertation, in progress).

2.2.1 Corpus. The corpus consisted of a 3-person discussion recorded in a soundproof room. The speakers were seated in a triangular arrangement and were filmed by two synchronized video cameras. The topic of the discussion was their current work, but some informal exchanges also took place. The total duration of the recording was about 20 minutes.

2.2.2 Data analysis. The films were coded by visual inspection using a U-matic videotape recorder. The gestures noted were head movements, hand movements, and direction of gaze. The vocal parameters considered were fundamental frequency, sound intensity, and duration. Back-channel signals produced by both listeners at the same time (double back-channel) and by only one of the listeners (single back-channel) were analyzed.

2.2.3 Results and comments. The number of back-channel signals produced varied across subjects. The results are presented in two parts: (1) a prosodic description and (2) a vocal and gestural typology.

The prosodic analysis dealt with the fundamental frequency (in Hz), the intensity (en dB), and the duration of the segment analyzed (in ms). For complex patterns (mixed rising and falling intonation contours), the largest variation was considered. Flat intensity curves were rare and were included in the falling patterns.

For the most part, the vocal back-channel signals exhibited a drop in intensity and frequency. The frequency variation was greater for falling patterns than for rising ones. The durations were more stable for listeners J and R than for listener I. This may be related to individual differences or to the particular roles played by the interlocutors in this conversation. In fact, J and R tended to speak to I, who appears to have been the favored partner for the other two. When double back-channel signals occurred, they had the same characteristics for both speakers.

The parameters used in defining a vocal and gestural typology were the variations in intensity and fundamental frequency for the vocal channel, and the direction of gaze and head and hand

movements for the gestural channel. For the hand, movements involving only one hand were distinguished from coordinated movements of both hands. In fact, there were no gestures of the left hand alone.

We can see that only a few of the potential combinations actually occurred. The most frequent patterns were those where a drop in intensity and frequency was associated with a change in direction of gaze, or with changes in head orientation and direction of gaze.

3. CONCLUSION

In this study, we showed how gestures can interact with vocal parameters in the production of conversational feedback. These preliminary results will be extended by further studies aimed at determining the precise temporal organization of the relationship between vocal and kinesic back-channel signals. One of these studies, based on the data obtained from a movement analyzer, is currently in progress.

REFERENCES

[1] BALLY C., 1925, *Le langage et la vie*, Librairie Droz, Librairie Giard.

[2] BERTRAND R., 1993, *Mise en relation de l'activité de la face avec les paramètres prosodiques et la chaîne segmentale dans un corpus d'interview télévisé*, Mémoire de DEA de phonétique expérimentale, fonctionnelle et appliquée, Université de Provence.

[3] BERTRAND R., 1994, *Approche pragmatique et prosodique de l'interaction conversationnelle*, Mémoire de D.E.A. de linguistique générale, Université de Provence.

[4] BIRDWHISTELL R., 1970, *Kinesics and context*, University of Pennsylvania Press.

[5] BOLINGER D., 1946, "Thoughts on Yep and Nope", *American Speech*, 21, 90-5.

[6] BOLINGER D., 1972, "Accent is predictable (if you're a mind-reader)", *Language*, 48, 633-44.

[7] BOLINGER D., 1985, *Intonation and its parts*, Edward Arnold.

[8] BOYER J., 1993, *Mise en relation des mouvements des bras et de la tête avec les paramètres prosodiques dans un corpus d'interview-reportage télévisé*, Mémoire de DEA, Université de Provence.

[9] CAVÉ C., GUAÏTELLA I. & SANTI, 1993, "Fréquence fondamentale et mouvements rapides des sourcils: une étude pilote", *Travaux de l'Institut de Phonétique d'Aix*, 15.

[10] CONDON W.S., 1976, "An analysis of behavioral organisation", *Sign Language Studies*, 13, 285-318.

[11] COSNIER J., 1988, "Grands tours et petits tours", in: Cosnier, Gelas & Kerbrat-Orecchioni (eds), *Echanges sur la conversation*, Editions du CNRS, Lyon, 175-84.

[12] FONAGY I., 1962, "Mimik auf glottaler Ebene" *Phonetica*, 8, 209-19.

[13] FONAGY I., 1967, "Hörbare Mimik", *Phonetica*, 16, 25-35.

[14] GUAÏTELLA I., 1991, *Rythme et parole: comparaison critique du rythme de la lecture oralisée et de la parole spontanée*, Thèse de Doctorat, Université de Provence.

[15] GUAÏTELLA I., 1995, "Mélodie du geste, mimique vocale", *Semiotica*, 103, 3/4.

[16] GUAÏTELLA I., CAVÉ C. & SANTI S., 1993, "Relations entre geste et voix: le cas des sourcils et de la fréquence fondamentale", *Actes du colloque Images et Langages, Multimodalité et modélisation cognitive*, Paris, 261-8.

[17] HEESE G., 1957, "Akzente und Betleitgeärden", Sprachforum, 2, 274-85.

[18] KERBRAT-ORECCHIONI C., 1991, *Les interactions verbales*, tomes 1 et 2, Colin.

[19] LAFOREST M., 1993, *Le back-channel en situation d'entrevue*, Recherches sociolinguistiques, 2, Université Laval, Québec.

[20] MEHRABIAN A., 1972, *Silent messages*, Wadsworth, Belmont.

[21] OHALA J.J., 1980, "The acoustic origin of the smile", *J. Acoust. Soc. Am.*, 68, 33.

[22] OHALA J.J., 1984, "An ethological perspective on common cross-language utilization of fo of voice", *Phonetica*, 41, 1, 1-16.

[23] VION R., 1992, *La communication verbale, Analyse des interactions*, Hachette.

[24] WERNER S., 1991, "Understanding 'hm', 'mhm', 'mmh'", *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix-en-Provence, vol.4, 446-8.

Table 1. Mean (M), maximum, and minimum variation in fundamental frequency, intensity, and duration for the different prosodic patterns. Single and double back-channel signals are shown for each speaker.

		↓F0. variation in Hz			↑F0. variation in Hz			↓Int. variation in db			↑ Int. variation in db			Duration in ms		
speaker	Case	M	Min.	Max.	M	Min.	Max.	M	Min.	Max.	M	Min.	Max.	M	Min	Max
single R	14	41	5	110	30	2	59	6 5	3	16	2	1	3	268 2	161	395
double R	12	26 9	6	63	41 5	24	59	5 5	1	11	6 3	4	8	252 2	167	472
single I	13	26 8	9	46	29 8	3	63	6 7	2	17	5 5	4	7	350	145	877
double I	1	33	33	33	/	/	/	/	/	/	/	/	/	157	157	157
single J	12	31 5	15	62	8 6	1	21	8 5	1	15	/	/	/	234	134	314
double J	11	12	2	33	15	1	38	6 7	2	14	5	5	5	264 5	173	444

Table 2. Number of cases of each intensity and frequency pattern.

Speaker	F0 ↘	F0 ↗	Complex F0	Int. ↘	Int. ↗	Complex Int.
Single R	9	3	2	9	2	3
Double R	9	1	2	8	1	3
Single I	5	4	4	9	1	3
Double I	1	0	0	1	0	0
Single J	9	3	0	10	0	2
Double J	8	3	0	9	1	1