IS IT EASIER TO LIPREAD ONE'S OWN SPEECH GESTURES THAN THOSE OF SOMEBODY ELSE ? IT SEEMS NOT!

Jean-Luc Schwartz & Christophe Savariaux

Institut de la Communication Parlée, Grenoble France
CNRS UMR 5009 – INPG – Université Stendhal
ICP, INPG, 46 Av. Félix Viallet, 38031 Grenoble Cedex 1, France - 04 76 57 47 12 - schwartz@icp.inpg.fr

ABSTRACT

In this paper, we attempt to adapt an experimental procedure inspired by Beardsworth and Buckner (1981), in which they studied the ability to recognise one's own versus somebody else's walking movements. They showed that certain subjects were better at recognising themselves than at recognising their friends, thanks to "some sort of kinesthetic-visual cross-modal transfer". We study the lipreading scores of French spoken digits uttered by 6 speakers and identified by the same 6 subjects. It appears that the performances are the same whether or not the subject is also the speaker. Hence we failed in our attempt to demonstrate a perceptuo-motor transfer in this experiment.

1. PERCEPTUO-MOTOR LINKS INSIDE AND OUTSIDE SPEECH

1.1. The "speech" debate

There is a classical debate in the field of speech communication about the nature of perceptual representations of speech gestures: purely sensory, and basically auditory, for tenants of auditory theories [1]; or purely motor for tenants of the motor theory [2]. Between these views of perception is one without action constraints, or perception without perceptual representations, as in the direct realist theory [3], we defend a view in which perception serves not only to understand gestures, but also to provide control signals to action: this is the Perception-for-Action-Control Theory [4]. This leads to an integrated sensori-motor framework in which perceptual and motor representations are acquired together and in interaction in the course of speech development, with a modelling approach based on the conceptual tools of "speech robotics" ([5], [6]).

1.2. Perceptuo-motor links outside speech

The nature of the perceptuo-motor links is of course also discussed outside the domain of speech communication. Since Johansson [7], there have been a large number of experimental psychology studies on the perception of biological movement, which provided the basis for more recent work on audiovisual speech perception with the same kinds of techniques [8]. The search of neural circuits in the cortex, providing possible links between perception and action, has received a recent strong impulse with the discovery of the so-called "mirror neurons" in the ventral premotor cortex of the monkey [9], responding to both the production and perception of complex actions. Two other centres seem to play an important part in the perception of biological movement in the monkey cortex: the superior temporal sulcus STS [10] and the parietal area AIP [11]. In the last few years, fMRI and PET neuroimagery linked these ideas to data about the human cortex. A recent meta-analysis of the functional anatomy of execution, mental simulation and observation of actions in the human cortex [12] confirms that there is a good overlap of cortical activation for these three types of tasks in the supplementary motor area SMA, the dorsal premotor cortex, the supramarginal gyrus and the superior parietal lobe. Grezes [13] summarises a number of experimental results on the observation of actions and defines a basic human cortical circuit, active specifically for the observation of actions but not their execution, consisting of visual areas in the occipital cortex (including MT/V5), the right superior temporal sulcus which could integrate information of the dorsal and ventral streams, and the left inferior parietal lobe. This picture is coherent with PET and fMRI data on lipreading (see a review in [14]), which add two important pieces: the activation of inferior prefrontal Broca's and Broca-homologue areas ([15], [16], [17]), and of the primary and secondary auditory cortex ([18], [19]. In summary, the cortical activation patterns of the lipreading system include inferior temporal, superior temporal, secondary and primary auditory cortex, supramarginal gyrus and inferior prefrontal regions.

1.3. The four levels of perceptuo-motor interactions

The nature of the perceptuo-motor links is well clarified by the distinction made by Viviani and Stucchi [20], between four possible levels of perceptuo-motor interactions: (i) the ecological level of perceptual covariations due to movement, (ii) active exploration, (iii) the generation of expectations linked to the efference copy mechanism, and (iv) a more abstract level in which procedural knowledge about general action principles would interfere with the elaboration of perceptual representations. It is to this level, according to the authors, that the motor theory of speech perception refers. To determine how real this fourth level is, one must find convincing experimental data, and Viviani and Stucchi suggest two directions. Firstly, their own work provides some of the most convincing (and quite rare) data on the influence of motor procedural knowledge on the perception of biological movement, in which they show that the "laws of human movement" could bias the perception of a given hand-drawn shape. In the field of speech perception, the search for experimental material able to display the influence of action on perception was at the core of the experimental research program at Haskins Labs. However, there does not seem to exist a clear-cut piece of data, though audio-visual speech perception data might fit well with this "fourth level" framework [21]. Secondly, Viviani and Stucchi mention a very interesting experiment. In this experiment, Beardsworth and Buckner [22] studied the ability to recognise one's own versus somebody else's movements from a recorded point-light display providing a schematic dynamic record of the walking movements. Subjects in the experiment were a group of college students who knew each other well, and it appears that certain subjects were better at recognising themselves than at recognising their friends. The crucial point in the interpretation is that they had never seen themselves walking from an external point of view, while they had seen their friends walking every day. Therefore, this result is interpreted by the authors as suggesting "some sort of kinesthetic-visual cross-modal transfer" (p. 19).

1.4. Proposal of a speech experiment

Beardworth and Buckner's experiment is rather simple and striking. Indeed, it seems to provide strong and direct evidence that action may guide or complete perception in some cases. When thinking about speech, it could be possible to apply it directly to a speaking face recognition study for example, but this is a bit marginal in respect to the core of the speech debate, centred on the identification of speech gestures rather than speakers. Hence we thought that it could be interesting to apply the same kind of idea to the problem of speech recognition. Of course, studying the *auditory* identification of one's own vs. somebody else's utterances does not seem very appealing, since one hears his/her own voice much more than other voices, even if the auditory pathway is partly different for one's own voice and for the other voices. But the *visual* identification should not suffer from the same objection: one almost never has the occasion to lipread his/her own gestures. Therefore we decided to prepare an experiment on the lipreading of self vs. other speech gestures, with the idea that if a benefit of seeing one's own gestures could be demonstrated, it would provide a strong cue in favour of the role of "motor procedural knowledge" in the elaboration of speech perceptual representation[1].

2. METHODOLOGY

2.1. Stimuli and subjects

For other purposes, we had video-recorded in our lab a series of isolated digits from 0 to 9 uttered in French by 10 speakers (5 male and 5 female) with 30 repetitions of each stimulus, hence an audio-visual base of 3000 stimuli that was used for audio-visual speech recognition experiments [23]. We decided to exploit this database. Stimuli had been recorded with the classical ICP experimental setup, including fixed head, excellent light conditions, and blue lips allowing automatic detection of lip parameters [24]. The image was centred on the lips, and these were the stimuli provided for lipreading identification. There is no reason to believe that the blue make-up was disturbing as it does not seem so different from classical make-up, and a number of perceptual experiments have already been performed on such kinds of stimuli. It could only help the subjects to focus on lip movements.

Since two of the five male speakers were informed of the experiment goal, they were discarded together with two female speakers in order to obtain an equal number of male and female speakers. Hence the experiment involved 6 speakers, 3 male and 3 female, none of whom were informed of the experiment's theoretical background. They were only told that systematic perception tests were being undertaken on all stimuli with all speakers to validate the audio-visual digit database they had recorded earlier, to be able to compare automatic recognition scores with perceptual data.

2.2. Organisation of the experimental task

Since the recorded material consisted of 6 tapes (one for each speaker), each containing 30 consecutive series of the 10 digits recorded in a random order. We decided to minimise the preparation time by exploiting the tapes directly, using two blocks of 5 consecutive series for each speaker. Hence altogether the experiment involved 600 stimuli: 10 repetitions of 10 digits by 6 speakers, grouped in 2 blocks of 50 stimuli per speaker, to be visually identified by the 6 subjects who were also the 6 speakers. A given subject identified a first series of the first block for each speaker, followed by a second series of the second block for each speaker.

The stimuli were presented on a monitor with no sound, and the subjects were instructed to visually

[1] To our knowledge, the question of possible differences between lipreading self and other gestures has not been studied yet.

identify each digit. Since the time between digits in the original tapes was quite small, the subject did not have enough time to write his response. Hence he was instructed to say the response in a loud voice. One of the authors was sat close to each subject and wrote the response on a prepared paper. In summary, the task consisted in identifying a digit and repeating it.

The next problem was to find a passage order correctly randomising the blocks between subjects, in order to take into account possible learning effects. For this aim, we defined a permutation order such that, for the first and second series, each speaker was respectively in 1st, 2nd, 3rd, 4th, 5th and 6th position for the 6 subjects, and also each case (Subject = Speaker) happened once in each position (see Table 1). Therefore, learning effects were randomised between subjects, between speakers and between *self* (Subject = Speaker) vs. *different* (Subject ≠ Speaker) conditions.

3. RESULTS

The results consist in 3600 responses provided by the 6 subjects to the 600 stimuli, and we grouped them in two ways: global confusion matrices between digits for all speakers and all subjects (Table 2) and global recognition scores for each subject and each speaker (Table 3). In this last case, we present separately the scores for the first and second block, in order to display possible learning effects. Let us analyse these results separately.

3.1. Confusion matrices between digits

Though this was not our primary focus, it is of interest to comment briefly on these confusion matrices, which show that there is a major confusion group including (1, 4, 5, 6, 7) in which the distinctions between (4, 5, 7) are quite difficult, 6 and at a lower level 1 being often confounded with these, though much better recognised. The mean global recognition score restricted to this group reaches only 57% (varying between 23% for 5 and 77% for 1). A second group including (0, 2, 3, 8, 9) contains stimuli that are almost always correctly identified by most subjects (except subject 3, see Section 3.2), and almost never confounded with stimuli in the first group. The mean global recognition score restricted to this group reaches 93%, and 95.5% for stimuli in the second blocks (see discussion of the learning effects in next section) and more than 99% for the three best subjects 1, 4 and 5 (see next section).

3.2. Analysis of global recognition scores

A two-way ANOVA was performed on global recognition scores, the two criteria being the 6 subjects and the 6 speakers (for this analysis we used the summed scores over the two blocks per speaker). It shows that the “speaker” factor, the “subject” factor and their interaction are significant.

Looking separately at each subject, it appears that indeed there is large inter-individual variability, with scores varying between 53% for subject 3 to 87% for subject 1. It appears also that for the best subjects; 1, 4 and 5, for whom the scores are high (from 79% to 87%), there is no difference between the first and the second blocks (no learning effect), while for subjects with lower scores there seems to be an effect, with a slight increase from the first to the second block for subjects 2 and 6 which enable them to reach the same scores as subjects 4 and 5 for the second block, and a very high increase from 47% to 60% for subject 3.

Last but not least, let us consider the “self” vs. “other” differences, summarised in Table 4. The difference is quite small, except for two subjects, subject 3 (56% vs. 52.6%) and subject 5 (90% vs. 77.2%). Globally, the mean “self” recognition is slightly higher than the recognition of “other” (77.5% vs. 74.7%) but the difference is not significant [Chi2(1) = 2.10, NS]. The case of subject 3 must be approached with great caution. Indeed, since he displays very large learning effects, the order of presentation might have played a large role in his results, and in any case the inter-block variability for this subject is so high that this result is not significant. The case of subject 5 is more interesting, since the self vs. other difference is very large, while the global scores are large with no learning effect. However, the “self” vs. “other” test performed on subject 5 alone compared to the 5 others does still not reach significance (Chi2(1) = 0.6, NS).

3.3. Discussion

Concentrating on our major goal, the result is negative since we observe no significant difference between the visual identification of self vs. other spoken digits. However, the case of subject 5 could not be discarded without further analysis. Indeed, the Beardsworth and Buckner study that provided us with our starting point reported very large inter-individual differences, with only some subjects better at recognising themselves than others. Though the present task is different, we decided to further explore the possibility that subject 5 was one of these subjects. This will be described in the next Section.

4. SECOND EXPERIMENT

4.1. Principle and methodology

In this experiment, we concentrated on three subjects, that is the two subjects who had provided some “self” vs. “other” gain, subjects 3 and 5, and a “reference subject”, subject 4, who appeared to behave globally more or less the same as subject 5 both as a subject (mean recognition scores for subjects 4 and 5 were 79.7 % and 79.3 %, respectively) and as a speaker (mean recognition scores for speakers 4 and 5 were 78.5% vs. 79.8%, see Table 3), except of course for the "self" vs. "other" difference, almost null for subject 4 (see Table 4). We also decided to concentrate on the group of the 5 most difficult digits (1, 4, 5, 6, 7), since the 5 others could not produce any difference but could blur an effect by saturating the results around high values. Previous Chi2 analyses on this restricted set showed that if the pattern of "self" vs. "other" results for subject 5 was not due to chance, and hence was stable for a new experiment, we needed 60 utterances per digit and per speaker (twice the available set, see Section 2.1) to be able to demonstrate a significant effect, instead of only 10 in the first experiment. We prepared a new tape with the extracted stimuli, containing 30 utterances of each of the 5 selected digits produced by the 3 selected speakers, hence 450 stimuli presented in a random order to eliminate the role of learning effects. This tape was presented twice to each subject, providing altogether 900 stimuli to each subject. Once more, the subject had to identify the visual stimuli with no sound, and pronounce the identified digit, which was written by the experimenter for further analysis.

4.2. Results

The results are displayed in Table 5. We did not analyse confusion matrices, since the pattern of errors should not be very different from the one in Exp. 1. Notice that some learning effects appeared in Exp. 2, for subject 3 (as in Exp. 1) but also for subject 4, though there was no learning for this subject in Exp. 1. The reason is almost certainly that the set of stimuli is more restricted in Exp. 2, with half as many speakers (3 instead of 6) and half as many digits (5 instead of 10), and much more utterances of a given digit by a given speaker (60 instead of 10). This is confirmed by the global lipreading scores, which were better in Exp. 2 than in Exp. 1 for the restricted digit set.

The pattern of results in Table 5 clearly discards our assumption presented in Section 3.3: Subject 5 does not confirm in Exp. 2 the benefit of "self" vs. "other" perception displayed in Exp. 1. Globally, the mean "self" recognition score averaged on the three subjects is 71.1%, vs. 68.4% for the "other" recognition score, and the difference is not significant (Chi2(1)=2.09, NS) [2]. Moreover, the pattern of between-subject differences is incoherent from one experiment to the other, Subject 4 displaying a "self" advantage in Exp. 2 but not in Exp. 1, while Subject 5 displays a "self" advantage in Exp. 1 but not in Exp. 2. Hence it appears in light of Exp. 2 that the case of Subject 5 in Exp. 1 was just a random fluctuation inside a global pattern in which lipreading produces basically the same performance for one's own and other's digits.

5. CONCLUSION

Our results are negative. We expected some difference between lipreading one's own and somebody else's speech gestures, and we did not find any significant difference. Of course, it is always embarrassing to comment on negative findings. It might be the case that there does exist some "self" vs. "other" difference but that we failed to find it because of a bad choice of experimental material, protocol, or subjects. Moreover, it is very important to notice that, while a positive finding would have been a piece of evidence in favour of perceptuo-motor links in the elaboration of perceptual representations, this negative finding does not demonstrate anything. Indeed, it could be the case that, even if action was involved in perception, it would not provide any gain to "self" perception, for example considering that a benefit of seeing one's own gestures and being able to replicate them more easily could be cancelled by a benefit of seeing somebody else visually well-known – and all our subjects knew each other.

In summary, the present story is just the story of a disappointment. But it was an occasion to attempt to better connect the perceptuo-motor literature with the speech literature, which is too seldom the case. Disappointments happen in life, and in sciences as well: we felt it worthwhile telling this disappointing story to our AVSP colleagues!

Acknowledgements

This work was supported by the French MENSR Cognitique project "ABISPA". Thanks to our subjects F. Sannier, C. Savariaux, L. Girin, A. Vilain, A. Amodio and P. Counillon. Thanks also to P. Teissier who helped us realise the first experiment, and to M.-A. Cathiard and S. Kandel for discussions in the preparation of this work.

[2] Even if we summarise the scores of Exp. 1 and 2, the "self" vs. "other" difference is still not significant (73.7% vs. 72.3%, Chi2(1) = 1.02, NS). But this summation is not acceptable, since the ratio of "self" and "other" tests is not the same in Exp. 1 and 2.

(First series)

Sub 1	**1**	2	3	4	5	6
Sub 2	3	4	5	6	1	**2**
Sub 3	2	**3**	6	1	4	5
Sub 4	6	5	**4**	2	3	1
Sub 5	4	6	1	**5**	2	3
Sub 6	5	1	2	3	**6**	4

(Second series)

3	2	**1**	5	6	4
1	3	4	**2**	5	6
2	4	5	6	**3**	1
4	5	6	1	2	3
6	1	2	3	4	**5**
5	**6**	3	4	1	2

Table 1 – Speaker permutation for each subject (one line per subject)

For example, the order for subject 3 in the first series was Spk 2, 3, 6, 1, 4 and 5. "Subject = Speaker" cases marked in grey

	1	2	3	4	5	6	7	8	9	0	other
1	278	0	4	31	4	0	2	0	1	0	40
2	1	310	2	0	0	1	0	10	5	6	25
3	0	0	336	1	0	0	0	7	0	1	15
4	1	0	1	193	16	6	137	0	1	0	5
5	3	1	1	61	81	17	172	1	0	1	22
6	7	1	0	12	12	266	25	1	0	2	34
7	1	0	0	61	40	23	214	2	0	0	19
8	0	0	5	0	0	0	0	351	0	1	3
9	0	1	0	0	0	0	0	1	344	0	14
0	0	2	1	0	3	0	1	2	9	333	9

Table 2 – Global confusion matrices between digits, averaged on all subjects and speakers

(uttered digits on lines, 360 answers per digit;; perceived digits on columns; "other" means other or no response)

	Sub 1			Sub 2			Sub 3			Sub 4			Sub 5			Sub 6			% per Spk
	1	*2*	Tot.	*1*	*2*	Tot.	*1*	*2*	Tot.	*1*	*2*	Tot.	*1*	*2*	Tot.	*1*	*2*	Tot.	
Spk 1	*86*	*86*	86	*78*	*72*	75	*38*	*56*	47	*86*	*80*	83	*80*	*76*	78	*74*	*80*	77	74.3
Spk 2	*92*	*92*	92	*72*	*82*	77	*28*	*42*	35	*74*	*70*	72	*82*	*70*	76	*70*	*78*	74	71.0
Spk 3	*86*	*88*	87	*74*	*80*	77	*46*	*66*	56	*88*	*80*	84	*78*	*74*	76	*76*	*72*	74	75.7
Spk 4	*90*	*82*	86	*76*	*82*	79	*66*	*64*	65	*82*	*80*	81	*78*	*84*	81	*84*	*74*	79	78.5
Spk 5	*86*	*88*	87	*70*	*84*	77	*60*	*68*	64	*80*	*86*	83	*86*	*94*	90	*68*	*88*	78	79.8
Spk 6	*84*	*80*	82	*72*	*70*	71	*44*	*60*	52	*70*	*80*	75	*72*	*78*	75	*74*	*76*	75	71.7
% per Sub	*87.3*	*86*	86.7	*73.7*	*78.3*	76	*47*	*59.3*	53.2	*80*	*79.3*	79.7	*79.3*	*79.3*	79.7	*74.3*	*78*	76.2	75.2

Table 3 – Global recognition scores for each subject, speaker and block (in %)

	Mean score	Self reco	Other reco
Sub 1	86.7 %	86 %	86.8 %
Sub 2	76 %	77 %	75.8 %
Sub 3	53.3 %	56 %	52.6 %
Sub 4	79.7 %	81 %	79.4 %
Sub 5	79.3 %	90 %	77.2 %
Sub 6	76.2 %	75 %	76.4 %
Mean	75.2 %	77.5 %	74.7 %

Table 4 – Mean, "self" and "other" recognition for each subject

	Sub 3	Sub 4	Sub 5	% per Spk
Spk 3	49.7	73.7	70.7	64.7
Spk 4	50	84.7	79.3	71.3
Spk 5	59.3	77.3	79	71.9
% per Sub	53.0	78.6	76.3	69.3

Table 5 – Global recognition scores for each subject and speaker in Exp. 2 (in %)

REFERENCES

1. Nearey, T.M. (1997). Speech perception as pattern recognition. *J. Acoust. Soc. Am., 101*, 3241-3254.

2. Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

3. Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *J. Phonetics, 14*, 3-28.

4. Schwartz, J.L., Abry, C., Boë, L.J., & Cathiard, M. (2001). Phonology in a theory of perception-for-action-control. In J. Durand & B. Laks (eds.) *Phonology: from Phonetics to Cognition.* Oxford: Oxford University Press (in press).

5. Schwartz, J.L., Boë, L.J., & Bessière, P. (submitted). Speech robotics, a computational framework for the study and modelling of speech acquisition and evolution. Submitted to *Artificial Intelligence Journal*.

6. Abry, C., & Badin, P. (1996). Speech Mapping as a framework for an integrated approach to the sensori-motor foundations of language. *4th Speech Production Seminar: from control strategies to acoustics*, 175-184, May 21-24, 1996, Autrans, France.

7. Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14*, 201-211.

8. Rosenblum, L.D., & Saldana, H.M. (1996). An audio-visual test of kinematic primitives for visual speech perception. *J. Exp. Psychology: Human Perception and Performance, 22*, 318-331.

9. Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions*. Cogn. Brain Res., 3*, 131-141.

10. Perrett, D.I., & Emery, N.J. (1994). Understanding the intentions of others from visual signals: neurophysiological evidence. *Curr. Psychol. Cognition, 13*, 683-694.

11. Taira, M., Mine, S., Georgopoulos, A.P., Murata, A., & Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Exp. Brain Res., 83*, 29-36.

12. Grezes, J., & Decety , J. (2001). Functional anatomy of execution, mental simulation, observation and verb generation of actions: a meta-analysis. *Human Brain Mapping*, in press.

13. Grezes, J. (2000). *Neuroanatomie fonctionnelle des liens entre la perception et l'action chez l'homme*. Thèse d'Université, Neuropsychologie, Université Claude Bernard, Lyon, France.

14. Campbell, R. (1998). How brains see speech : the cortical localisation of speechreading in hearing people. In R. Campbell, B. Dodd & D. Burnham (eds.) Hearing by eye, II. Perspectives and directions in research on audiovisual aspects of language processing (pp. 177-193). Hove (UK): Psychology Press.

15. Blasi, V., Paulesu, E., Mantovani, F., Menoncello, L., De Giovanni, U., Sensolo, S., Perani, D., & Fazio, F. (1999). Ventral prefrontal areas specialised for lip-reading: a PET activation study. Human Brain Mapping.

16. Surguladze, S., Calvert, G., Brammer., M., Bullmore, E.S, Campbell, R., Williams, S.C.R., & David, A.S. (1999). Auditory-visual speech integration in schizophrenia: an fMRI study. *Human Brain Mapping*

17. MacSweeney, M., Amaro, E., Calvert, G.A., Campbell, R., David, A.S., McGuire, P.K., Williams, S.C.R., Woll, B., & Brammer, M.J. (2000). Activation of auditory cortex by silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport*, *11*, 1729-34.

18. Calvert, G.A., Bullmore, E., Brammer, M., Campbell, R., Woodruff, P., McGuire, P., Williams, S.C.R., Iversen, S.D., & David, A.S. (1997). Activation of auditory cortex during silent speechreading. *Science*, *276*, 593-596.

19. Campbell, R., Calvert, G., Brammer, M., MacSweeney, M., Surguladze, S., McGuire, P., Woll, B., Williams, S., Amaro, E., & David, A.S. (1999). Activation in auditory cortex by speechreading in hearing people: fMRI studies. *Proc. AVSP'99*.

20. Viviani, P., & Stucchi, N. (1992). Biological movements look uniform: evidence of motor-perceptual interactions. *J. Exp. Psychology: Human Perception and Performanc*, *18*, 603-623.

21. Schwartz, J.L., Robert-Ribes, J., & Escudier, P. (1998). Ten years after Summerfield ... a taxonomy of models for audiovisual fusion in speech perception. In R. Campbell, B. Dodd & D. Burnham (eds.) *Hearing by eye, II.* (pp. 85-108). Hove (UK): Psychology Press.

22. Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seing one's body. *Bulletin of the Psychonomic Society, 18*, 19-22.

23. Teissier, P., Robert-Ribes, J., Schwartz, J.L., & Guérin-Dugué, A. (1999). Comparing models for audiovisual fusion in a noisy-vowel recognition task. *IEEE Trans. Speech and Audio Processing*, 7, 629-642.

24. Lallouache, M.T. (1990). Un poste 'visage-parole'. Acquisition et traitement de contours labiaux. Proceedings of the XVIII^{èmes} Journées d'Études sur la Parole, Montréal, pp. 282-286.