



Neuropsychological *Trends*

3
April 2008

Simona Amenta - Michela Balconi

Understanding irony: an ERP analysis on the elaboration
of acoustic ironic statements 7

Frédéric Joassin - Pierre Maurage - Salvatore Campanella

Perceptual complexity of faces and voices modulates
cross-modal behavioral facilitation effects 29

Joy Helena Wymer B.S.

Psychological and neuropsychological correlates
of postconcussional disorder 45



Perceptual complexity of faces and voices modulates cross-modal behavioral facilitation effects

Frédéric Joassin¹ - Pierre Maurage¹ - Salvatore Campanella²

¹ Cognitive Neurosciences in Clinical Psychology Research Units, Department of Psychology, Catholic University of Louvain, Louvain-la-Neuve, Belgium

² Department of Psychiatry, Brugmann Hospital, Free University of Brussels, Brussels, Belgium

frederic.joassin@psp.ucl.ac.be

ABSTRACT

Joassin et al. (Neuroscience Letters, 2004, 369, 132-137) observed that the recognition of face-voice associations led to an interference effect, i.e. to decreased performances relative to the recognition of faces presented in isolation. In the present experiment, we tested the hypothesis that this interference effect could be due to the fact that voices were more difficult to recognize than faces. For this purpose, we modified some faces by morphing to make them as difficult to recognize as the voices. Twenty one healthy volunteers performed a recognition task of previously learned face-voice associations in 5 conditions: voices (A), natural faces (V), morphed faces (V30), voice-natural face associations (AV) and voice-morphed faces associations (AV30). As expected, AV led to interference, as it was less well and slower performed than V. However, when faces were as difficult to recognize as voices, their simultaneous presentation produced a clear facilitation, AV30 being significantly better and faster performed than A and V30. These results demonstrate that matching or not the perceptual complexity of the unimodal stimuli modulates the potential cross-modal gains of the bimodal situations.

Keywords: Cross-modal; Facilitation; Faces; Voices

1. INTRODUCTION

In our everyday life, we are constantly required to associate pieces of information arising from distinct sensory modalities, in order to maintain a coherent

and unified representation of the world. We live in a multimodal world, and this is particularly true for our social interactions. For instance, to know who is speaking, we need to associate distinct visual and auditory percepts such as heard syllables and lips movements (Calvert et al., 2000). But to know who we are speaking to, our memory of people needs also to encode, store and retrieve multimodal representations of familiars, based on the distinct representations of their faces, voices, names, semantics and so on (Paller et al., 2003).

Over the last decade, more and more studies have been devoted to the study of the multimodal processing of objects, and they have brought to light two major behavioral effects: facilitation versus interference effects. Facilitation is characterized by a significant gain (in terms of shorter correct latency and/or better performances) for congruent bimodal stimulations relative to unimodal (visual or auditory) ones (Miller, 1982; Hughes et al., 1994; Frens et al., 1995). Two main types of cognitive models have been proposed to explain this facilitation effect: on the one hand the “race models” (Downar et al., 2000; Raab, 1962; Murray et al., 2001) postulating an independent and parallel processing of the different stimuli, without interaction between modalities, and explaining the faster RT’s observed in crossmodal condition as a simple consequence of the competition between modalities, the fastest processed stimulus mediating the response. On the other hand, the co-activation models (Miller, 1982; Schröger & Widmann, 1998) which explain the facilitation effect on the basis of an interaction between modalities, postulating either only a late interaction after modality-specific processing (“independent coactivation model”) or early integration and mutual influence between stimuli from the beginning of the process (“interactive coactivation model”).

Nevertheless, cross-modal interactions can also lead to interference, that is to say, a deteriorated or biased performance in bimodal situations relative to unimodal ones. In the visuo-auditory domain, two striking examples of such interactions are the McGurk (McGurk & McDonald, 1976) and the ventriloquist effects (e.g. Alais & Burr, 2004). In the former one, the simultaneous presentation of a syllable (e.g. /ba/) and an incongruent lip movement (e.g. /ga/) leads to the erroneous perception of the syllable /da/. In the latter one, the origin of a sound is attributed by mistake to the dummy and not to the artist’s mouth. These two examples show that vision can bias audition. But conversely, audition can also alter vision, as Sekuler et al. (1997) have shown that sounds can bias the perception of moving visual targets. Shams et al. (2002) have also recently demonstrated that the synchronous presentation of multiple auditory beeps and a single visual flash induces the subjective perception of multiple flashes, proving that the alteration of vision by audition and vice versa are not limited to the cases in which the pieces of information are ambiguous or moving.

Moreover, interference is not limited to perception but also occurs at later stages of the information processing stream. Indeed, Joassin et al. (2004), using human faces and voices, i.e. highly ecological and complex stimulations, examined cross-modal visuo-auditory interactions in a recognition task of previously learned face-voice associations. The recognition task was carried out in three conditions: either voices alone, either faces alone, either the simultaneous and congruent presentation of a face and a voice. The authors observed that the bimodal condition was significantly slower performed than the visual condition, i.e. that face-voice associations were more slowly recognized than faces alone. They interpreted this result as an interference of audition on vision, as the additional auditory information slowed down the processing of the visual information. This raised the question as to why the recognition of voices hampered rather than helped the recognition of faces in this case. One potential explanation is that voices and faces do not share the same degree of difficulty of recognition, or, in other words, that we could be more expert in face recognition than in voice recognition. Indeed, Joassin et al. (2004) observed that voices were recognized significantly more slowly than faces, and it has already been observed that voices are globally more difficult to recognize than faces (Schweinberger et al., 1997; Ellis et al., 1997). More specifically, Hanley et al. (1998) showed that it is more difficult to retrieve biographical information about famous people on the basis of their voice rather than on their face. Voices often produced a feeling of familiarity only. In a further study, Hanley and Turner (2000) hypothesized that this familiarity-only effect of voices could be explained by a general lower level of familiarity for voices than for faces, making the retrieval of biographical information from voices more difficult. Thus, they tried to make the faces as difficult to recognize as voices, by presenting the faces of famous people out of focus. Their results indicated that, under these conditions, faces and voices produced identical familiarity-only effects and that biographical information was as difficult to retrieve after the presentation of both faces and voices. For the authors, when steps are taken to reduce familiarity level for faces to a level equivalent to that for voices, faces and voices behave in exactly the same way. Moreover, a simulation with the IAC (Interactive Activation and Competition) model of Burton et al. (1999) confirmed these results and showed that they can be explained by weaker connections between VRU (Voice Recognition Units) and PIN (Person Identity Nodes) than between FRU (Face Recognition Units) and PIN.

In this context, it is possible that voices, being more difficult to recognize than faces, slowed down the recognition of face-voice associations in comparison with the recognition of faces presented in isolation. Following Hanley and Turner (2000), the purpose of the present experiment was thus

to put this explanation to the test, with the hypothesis that, if the bimodal condition relied on simultaneous presentation of faces and voices that are equally difficult to recognize, it should facilitate rather than hamper recognition, i.e. that such face-voice associations should be recognized faster and better than faces and voices presented in isolation. To examine this question, we attempted to match the levels of recognition of faces and voices by reducing the level of performance for faces. We increased the difficulty of recognition of the faces by using a morphing software, so that they led to recognition performances equivalent to those of voices, and we compared the recognition of face-voice associations in which faces were either easier or as difficult to recognize as voices.

2. METHODS

2.1. *Participants*

Twenty-one healthy volunteers (10 females, mean age: 21.8, SD: 4.45) took part to this experiment. All but three were right-handed and all had normal vision and audition.

2.2. *Stimuli*

Eight associations (4 females) between a face, a voice and a Belgian family name were created to form 8 “schematic” people. The faces were black-and-white pictures of unknown individuals selected from the Stirling Face Database (<http://www.pics.psych.stir.ac.uk>). All faces were presented in frontal position with a neutral expression. They were downloaded onto a Macintosh computer and were edited by Adobe Photoshop 5.0 (Adobe Systems Incorporated). Gray-scale images were created and scaled to 274 x 350 pixels (123 x 96 mm, corresponding to a visual angle of $7.04^\circ \times 5.5^\circ$, Figure 1).

The family names were chosen from the “Belgian National Institute of Statistics” so that each name had nearly the same frequency of occurrence in the Belgian population and was constituted by 6 letters and 2 syllables.

The voices were numeric recordings of unknown males and females saying the French word “bonjour” (“hello”). Intensity and duration (700 ms) were normalized with Goldwave® (Gold Wave Inc.).



Figure 1. Examples of faces with their percentages of morphing

Three learning sessions, performed on three consecutive days, served to familiarize participants with these associations (first day: female associations, second day: male associations, third day: male and female associations). The encoding of the associations was carried out by a computer presentation using Superlab 6.1 software (Cedrus Corporation). Each association was presented separately and participants were asked to try to remember the family name, with no time pressure. The encoding was repeated on request. To ensure that the associations were correctly encoded, several recognition and identification tests were performed: face-name matching, voice-face matching, voice-name matching, recognition of the associations among distractors (unknown and erroneous associations), recall of the name of each face and each voice. If errors occurred, they were directly corrected and a new encoding was performed. Each learning session was continued until accuracy reached 100% on each test. The experimental session took place on the fourth day.

In order to make the faces as difficult to recognize as the voices during the experimental session, they were modified by morphing using MorphTM 2.5 (Gryphon software corporation). Firstly, we tried to find which proportion of morphing would make the faces more difficult to recognize than the natural (that is non-morphed) faces. For this purpose, we carried out a pre-test on 12 healthy participants (mean age: 18.9 SD: 0.49) who did not take part in the main experiment. In this pre-test, each face was mixed with two other faces of the same gender in 5 different proportions: natural, 90% of a face – 10% of another face of the set (morphing at 10%), 80-20% (morphing at 20%), 70-30% (morphing at 30%) and 60-40% (morphing at 40%). We

thus obtained 80 different faces (5 proportions x 8 faces x 2 morphes). After the learning phase, participants were asked to categorize each presented stimulus (natural or morphed faces) according to its family name (e.g. “is it Detiez or Goffin?”), by pressing one of two keys on a stimpad. Each stimulus was presented for 700 ms in 8 blocks each containing only two identities. Each stimulus was presented twice in each block. Accuracy and latencies were recorded using e-prime (Schneider et al., 2002), and the statistical analyses were performed only on the correct answers. The results showed that recognition of blended faces at 10% and 20% did not differ significantly from that of natural faces, neither on latencies nor on accuracy. However, faces morphed at 30% (RT: $F(1,11) = 49.64$, $P < 0.001$; accuracy: $F(1,11) = 35.00$, $P < 0.001$) and 40% (RT: $F(1,11) = 52.32$, $P < 0.001$; accuracy: $F(1,11) = 126.90$, $P < 0.001$) were identified significantly more slowly and less accurately than unmorphed faces. Morphing was thus a relevant method to increase the difficulty of recognition of the faces. For the main experiment, we used natural faces and faces morphed at 30% (Figure 1), as this proportion of morphing was sufficient to significantly slow down the recognition without impairing seriously its accuracy (mean % of correct responses: natural faces = 97.2%; 30% morphed faces = 90.9%; 40% morphed faces = 78.4%).

2.3. Procedure

The 21 participants (see description above) of the main experiment were first trained to recognize each association (same procedure as described above), i.e. to retrieve the name linked to each association. Please note that the training was performed on the natural faces. On the fourth day, they performed the experiment, in which they were confronted with 5 different conditions: voices alone (auditory condition A), natural faces (visual condition V), faces morphed at 30% (V30), simultaneous presentation of voices and natural faces (auditory-visual condition AV), simultaneous presentation of voices and 30% morphed faces (AV30). Sixteen blocks of trials were presented, each block containing only two different identities that changed across blocks. Within each block, 18 trials were randomly presented (6 A trials, 6 visual trials including 3 V and 3 V30, and 6 auditory-visual trials including 3 AV and 3 AV30). Half of the trials corresponded to one identity, the other half to the other identity. Each trial was formed with a fixation cross appearing for 300 ms at the center of the screen, followed by the stimulus appearing for 700 ms, and an empty intertrial of 1200 ms.

The participants were instructed to categorize as fast and accurately as possible each stimulus (face, voice or association) according to its identity among two, which appeared on the screen at the beginning of each block. They had to answer by pressing one of two keys on the keyboard. They were also informed that there were no incongruent trials, i.e. bimodal situations in which faces and voices did not share the same identity. Participants were not informed that half of the faces that they would see had been modified by morphing.

Participants seated in a quiet room, with their head at approximately 50 cm away from the computer screen. Stimulus presentation and data recordings were provided by e-prime (Schneider et al., 2002) implemented on a Dell laptop. Voices were presented through headphones and participants were asked to adjust the volume to insure a correct and comfortable perception.

2.4. Statistical analyses

The statistical analyses were performed on the mean latencies and percentages of correct responses. For accuracy, we did not include errors, omissions or responses slower than 1000 ms.

We performed firstly two repeated measures ANOVAs with the conditions (A, V, AV, V30 and AV30) as the within factor, on latencies and accuracy separately. Subsequently, two distinct set of analyses, depending on the kind of faces (natural vs. morphed) were performed. Two ANOVAs with modality as within factor (auditory, visual, auditory-visual) were carried out on latencies and accuracy of the trials in which natural faces were presented (A, V and AV). The same ANOVAs were performed on the trial containing morphed faces (A, V30 and AV30). When appropriate, subsequent paired t-tests comparing each conditions to the other two (A vs. V, A vs. AV, V vs. AV, A vs. V30, A vs. AV30, V vs. AV30) were used. We also compared V and V30, AV30 and AV, and AV30 and V in 3 separate one-way ANOVAs. Were considered as statistically significant the F or t values whose $p \leq 0.05$. Greenhouse-Geisser corrections were applied when appropriate.

3. RESULTS

Mean latencies and percentages of correct responses for each condition are shown in Table 1 (standard deviations between parentheses).

Table 1. Mean reaction times (in milliseconds) and percentages of correct responses (standard deviations in parentheses) of the five conditions: A = auditory (voices), V = Visual (natural faces), V30 = visual (faces morphed at 30%), AV = auditory-visual (voice-natural face associations), AV30 = auditory-visual (voice-morphed face associations)

	A	V	AV	V30	AV30
RT (ms)	655.5 (121.93)	562 (97.14)	568.5 (98.84)	642.05 (119.98)	607.44 (102.77)
%	87.3 (9.7)	93.8 (6.72)	90.6 (6.1)	85.21 (9.25)	91.17 (7.63)

3.1. Comparison of the 5 conditions

The ANOVA performed on latencies and accuracy showed a significant effect (latencies: $F(4,80) = 40.01$, $p < 0.001$; accuracy: $F(4,80) = 12.19$, $p < 0.001$), indicating significant differences between the conditions.

Comparison of A, V and AV

The ANOVA performed on percentages of correct responses revealed a significant effect ($F(2,40) = 11.39$, $p < 0.001$). The subsequent paired t-tests showed that A was significantly less well performed than V ($t(20) = -5.08$, $p < 0.001$) and AV ($t(20) = -2.31$, $p < 0.03$). AV was also less well performed than the V ($t(20) = -2.33$, $p < 0.03$, Figure 2, left lower part).

The ANOVA performed on latencies showed a significant effect ($F(2,40) = 57.71$, $p < 0.001$). Subsequent paired t-tests showed that V was performed significantly more quickly than A ($t(20) = -9.38$, $p < 0.001$) and AV ($t(20) = -3.14$, $p < 0.01$). AV was also significantly faster performed than A ($t(20) = -7.51$, $p < 0.001$, Figure 2, left upper part).

These results indicate that natural faces were easier to recognize than voices, as they were faster and better recognized than voices, and that the simultaneous congruent presentation of both pieces of information disturbed recognition. Like Joassin et al. (2004), we are thus here in the case of an interference effect of audition on vision because the present results showed that adding an auditory information to the faces slowed down and impaired the recognition of these faces.

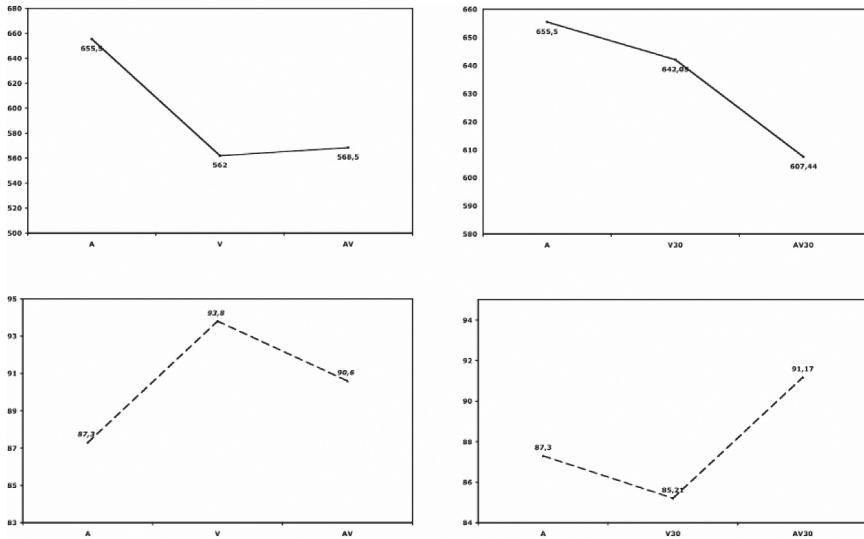


Figure 2. Mean latencies (upper part) and percentages of correct responses (lower part) for auditory, visual and auditory-visual conditions, according to the kind of faces: natural (left) or morphed (right)

Comparison of A, V30 and AV30

The ANOVA performed on accuracy revealed a main effect of the modality ($F(2,40) = 17.72$, $p < 0.001$). Paired student t-tests showed that voices and morphed faces were recognized in the same latencies ($t(20) = -2.03$, n.s) but that morphed faces-voices associations were recognized significantly quicker than either voices ($t(20) = -5.49$, $p < 0.001$) either morphed faces alone ($t(20) = -3.70$, $p < 0.001$, Figure 2, right lower part).

The ANOVA carried out on accuracy showed the same results, i.e. a main effect of the modality ($F(2,40) = 11.30$, $p < 0.001$), examined by subsequent paired Student t-tests showing that A and V30 did not differ significantly ($t(20) = 1.42$, n.s), but that AV30 was significantly better performed than A ($t(20) = 3.34$, $p < 0.001$) and V30 ($t(20) = 5.22$, $p < 0.001$, Figure 2, right upper part).

These results clearly showed that (1) modifying faces by morphing made them as difficult to recognize as the voices, and (2) the simultaneous presentation of congruent faces and voices identically difficult to recognize facilitated the recognition process, as morphed face-voice associations were

significantly better and quicker recognized than either faces either voices presented in isolation.

Comparisons of natural vs. morphed faces and associations

In this section, we directly compared the correct latency and accuracy elicited by natural vs. morphed faces, (1) to ensure that the morphing procedure made the faces significantly more difficult to recognize and (2) to examine the extent of the facilitation observed in AV30.

Firstly, ANOVAs with condition (V and V30) as within factor were performed. As expected, they showed that natural faces were quicker ($F(1,20) = 64.86$, $p < 0.001$) and better ($F(1,20) = 22.99$, $p < 0.001$) recognized than morphed faces.

Secondly, we compared AV30 and AV, i.e. a bimodal condition using faces difficult to recognize but eliciting facilitation and a bimodal condition using faces easy to recognize but failing to produce facilitation. On latencies of responses, the ANOVA revealed that AV was performed significantly faster than AV30 ($F(1,20) = 12.79$, $p < 0.001$), but on accuracy, it failed to show any significant difference between both conditions ($F(1,20) = 0.22$, n.s).

Thirdly, we compared AV30 to V, which was the easiest condition of all (better and quicker performed than A, V30 and AV). The ANOVA performed on latencies showed that natural faces were recognized significantly faster than the morphed face-voice associations ($F(1,20) = 41.90$, $p < 0.001$). But on accuracy, the ANOVA did not reveal any significant difference between both conditions ($F(1,20) = 3.72$, n.s).

This last set of results obviously showed that the morphing technique produced blended faces that were significantly more difficult to recognize than the natural ones and, more importantly, that adding a voice to such morphed faces facilitated so much their recognition that the performances – at least in terms of accuracy – became similar to the performances obtained with the easiest stimuli to be recognized, i.e. the natural faces.

4. DISCUSSION

The purpose of the present experiment was to compare the recognition of face-voice associations in which faces were either easier or as difficult to recognize as voices, i.e. when faces and voices were matched or not for recognition performance, with the hypothesis that the simultaneous presentation

of faces and voices equally difficult to recognize should facilitate rather than hamper recognition. To examine this question, we increased the difficulty of recognition of the faces in using a morphing procedure.

Firstly, the present results clearly show that the morphing technique that we used was reliable. It provided faces which were significantly more difficult to recognize than non-morphed faces, as indexed by significantly longer reaction times and decreased accuracy. It must be noted however that this increase of difficulty did not prevent the recognition of the faces, the mean percentage of correct responses remaining largely over random. Moreover, morphing allowed us to modify the faces so that they became as difficult to recognize as the voices. The two goals of the morphing procedure were thus fully achieved.

Secondly, our results confirm the observations of Joassin et al. (2004) that, when faces are easier to recognize than voices, their simultaneous presentation produce interference. They also confirm our main hypothesis that associations of auditory and visual information of equal complexity should elicit facilitation rather than interference. Actually, when faces were modified to become as difficult to recognize as voices, their simultaneous presentation facilitated recognition. This facilitation was observed on both reaction times and accuracy. It seems thus that adding an auditory information can help the recognition when faces are difficult to recognize by themselves. Moreover, this facilitation was enough important to bring AV30 at the same level of accuracy than the easiest unimodal condition of all, i.e. the recognition of natural faces. Nevertheless, we must acknowledge that this facilitation remains a relative facilitation and not an absolute facilitation. AV30 speeded up the responses, but only in comparison with A and V30, V and AV remaining faster performed than AV30. It could be due to the fact that in AV, the recognition was mainly driven by the processing of the natural faces and that a processing of the visual information was sufficient to take a correct decision. Such attentional bias towards a dominant modality has already been observed (Bushara et al., 2003), notably on the visuo-auditive categorization of emotional stimuli (Ethofer et al., 2006). In this case, the auditory information could have parasitized the processing of faces, leading to a decreased performance in AV than in V.

On the other hand, in the case of associations between voices and morphed faces, neither stimuli was easier to identify than the other and there was no reason for an attentional bias towards faces or voices to occur. In this case, the recognition was based on the processing of both faces and voices, creating a facilitation relative to morphed faces or voices presented in isolation. But the conjunction of morphed faces and voices harder to recognize than natural faces was probably not sufficient to speed up recognition at the

level of natural face recognition. Further experiments, using electrophysiological and neuroimaging techniques, could help to better understand the cognitive and cerebral mechanisms involved in the visuo-auditory interactions between faces and voices.

Joassin et al. (2004) postulated that their observed interference effect could be due to a temporal asynchrony in the quantity of information available for recognition, all the visual information needed to recognize a face being available at the onset while the auditory information contained in voices was minimal at the onset and increased with time. However, in the present experiment, as this potential temporal asynchrony was present in both kinds of associations, it is obviously not sufficient to explain the different effects observed. The present data indicate rather that cross-modal facilitation or interference effects depend on the respective levels of difficulty of the unimodal information that are associated in the bimodal condition. In the present case, it is only when the visual stimuli matched the auditory stimuli in difficulty, reflected by equivalent behavioral performance, that we could observe a clear facilitation.

Cross-modal literature has demonstrated that there are several factors that determine the potential gain elicited by bimodal congruent stimulations, notably their proximity in time and space (Wallace et al., 1996; Meredith & Stein, 1986; Radeau, 1994) and their semantic congruency (Stein et al., 1989; Sekuler et al., 1997; Calvert et al., 2001). The present results add another factor of crossmodal gain as they demonstrate that the perceptive complexity of the stimulations presented in distinct sensory modalities influence their integration and the associated behavioral effects.

Another cross-modal factor could explain our results. Indeed, they can be related to the principle of inverse effectiveness proposed, at the neuronal level, by Meredith and Stein (1986). In this principle, the neuronal responses in bimodal conditions are maximal when the contributing unimodal stimuli are minimally effective. Our results are in accordance with inverse effectiveness and could be the transposition, at a cognitive level, of this principle as we observed that the cross-modal enhancement was maximal when both faces and voices were equally difficult to recognize, i.e. were both minimally effective for the recognition task to perform.

In the present experiment, we increased the perceptive difficulty of the visual stimuli and we observed the effects of such a manipulation on a recognition task. We cannot thus define which level of the processing stream was more influenced by this matching of the perceptive difficulty. Further experiments are needed to clarify this point. However, we think that these findings are of particular relevance for prosopagnosia, as it has been shown that brain-damaged patients rely on voices to identify friends and relatives, either overt-

ly (Pallis, 1955) or covertly (de Haan et al., 1987). Our results, showing that voices help recognition when face recognition is made more difficult, are, to our knowledge, the first to replicate these observations on healthy subjects. Moreover, some cases of patients with slowly progressive impairment of their recognition of familiar people from voices and faces have also been described (Gentileschi et al., 1999; Gainotti et al., 2003). Overall, these data on brain-damaged patients and healthy subjects are in line with the hypothesis that some people recognition disorders could be due to an impaired access from unimodal channels to a multimodal person-recognition system (Gainotti et al., 2003) and that some convergence point between different modes of recognition are needed to access semantics and to identify individuals efficiently.

In conclusion, this study showed that faces are easier to recognize than voices because: (1) natural faces were significantly easier to recognize than voices, supporting previous observations (Ellis, 1997; Schweinberger et al., 1997); (2) we were brought to increase the perceptive complexity of the faces to observe a facilitation effect when presented in association with the voices. In the same way, Hanley et al. (1998) observed that it is more difficult to retrieve biographical information from a voice than from a face. Hanley and Turner (2000) showed that, when faces were made as difficult to recognize as voices, biographical information were as difficult to retrieve after the presentation of both stimuli, which produced identical familiarity-only feelings. The authors postulated that voices are associated with lower overall levels of familiarity than faces. A simulation with the IAC model of Burton et al. (1999) confirmed this interpretation in showing that their results could be explained by weaker connections between VRU and PIN than between FRU and PIN. On the whole, this body of data lead us to think that we are more expert in face recognition than in voice recognition. The question that arises now is why, and the answer should maybe be searched in the evolution and phylogenesis of mankind. Indeed, it is recently that the need to recognize and identify people on the basis of their voice only has really developed. This ability was probably less crucial before the invention and the wide diffusion of modern communication technologies such as the telephone or the radio, as in most cases the people who had to be recognized were potentially visible. It is thus an exciting idea to think that, if this hypothesis is correct and if the communication technologies continue to call for auditory expertise, the superiority of vision on audition in the identification of people could be a transient effect and that, in a long-term evolutionary perspective, we could be brought to develop our ability to identify voices as easily as faces.

ACKNOWLEDGMENTS

The first and second authors are supported by the National Fund for Scientific Research (FNRS, Belgium).

We gratefully thank Prof. Raymond Bruyer for his helpful comments.

REFERENCES

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optical bimodal integration. *Current Biology*, 14, 257-262.
- Burton, A.M., Bruce, V., & Hancock, P.J.B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, 23, 1-31.
- Bushara, K.O., Hanakawa, T., Immish, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience*, 6, 190-195.
- Calvert, G.A., Campbell, R., & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Current Biology*, 10, 649-657.
- Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*, 14, 427-438.
- de Haan, E.H.F., Young, A.W., & Newcombe, F. (1987). Face recognition without awareness. *Cognitive Neuropsychology*, 4, 385-415.
- Downar, J., Crawley, A.P., Mikulis, D.J., & Davis, K.D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, 3, 277-283.
- Ellis, H.D., Jones, D.M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology*, 88, 143-156.
- Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., Reiterer, S., Grodd, W., & Wildgruber, D. (2006). Impact of voice on emotional judgment of faces: an event-related fMRI study. *Human Brain Mapping*, article online in advance of print.
- Frens, M.A., Van Opstal, O.A., & Van der Willigen, W.R. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception and Psychophysics*, 57, 802-816.
- Gainotti, G., Barbier, A., & Marra, C. (2003). Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain*, 126, 792-803.

- Gentileschi, V., Sperber, S., & Spinnler, H. (1999). Progressive defective recognition of familiar people. *Neurocase*, 5, 407-424.
- Hanley, J.R., Smith, S.T., & Hadfield, J. (1998). I recognise you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology*, 51A, 179-195.
- Hanley, J.R., & Turner, J.M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology*, 53A, 1105-1116.
- Hughes, H.C., Reuter, L.P., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology Human Perception and Performances*, 20, 131-153.
- Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M. & Campanella, S. (2004). When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. *Neuroscience Letters*, 369, 132-137.
- McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Meredith, M.A., & Stein, B.E. (1986). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, 75, 1843-1857.
- Miller, J.O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14, 247-279.
- Murray, M.M., Foxe, J.J., Higgins, B.A., Javitt, D.C., & Schroeder, C.E. (2001). Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia*, 39, 828-844.
- Paller, K.A., Ranganath, C., Gonsalves, B., LaBar, K., Parrish, T.B., Gitelman, D.R., Mesulam, M.M., & Reber, P.J. (2003). Neural correlates of person recognition. *Learning & Memory*, 10, 253-260.
- Pallis, C.A. (1955). Impaired identification of faces and places with agnosia for colors. *Journal of Neurology, Neurosurgery and Psychiatry*, 18, 218-224.
- Raab, D.H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24, 574-590.
- Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Current Psychology and Cognition*, 13, 3-51.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.
- Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, 35, 755-759.

- Schweinberger, S.R., Herholz, A., & Stief, V. (1997). Auditory long-term memory: repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology*, 50 (A), 498-517.
- Sekuler, R., Sekuler, A.B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385, 308.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147-152.
- Stein, B.E., Meredith, M.A., Huneycutt, W.S., & McDade, L. (1989). Behavioural indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, 1, 12-24.
- Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, 76, 1246-1266.