Recalling Semantic and Episodic Information From Faces and Voices: A Face Advantage

Serge Brédart and Catherine Barsics
University of Liège

Abstract

Recent studies have demonstrated that the retrieval of biographical information about familiar people is easier when we see their faces than when we hear their voices. This advantage of faces over voices has been observed for the retrieval of semantic information (e.g., a person's occupation) as well as for the recall of episodic information (e.g., specific memories associated with a person). In this article, we outline a recent progression of studies that have demonstrated this advantage of faces over voices by comparing the retrieval of semantic and episodic information following person recognition from faces and from voices. We show that the face advantage is a robust phenomenon that persists whatever the type of target persons (celebrities, personally familiar people, or newly learned persons).

Keywords

The retrieval of information about familiar people following the recognition of their face has been extensively investigated during the 25 years following the publication of Bruce and Young's (1986) seminal model of face processing. According to this model, there are three sequential stages involved in the recognition of a familiar face. First, a sense of familiarity is associated with the seen face. Following this first stage, the beholder may retrieve identity-specific semantic information (e.g., the target's person occupation or nationality) or episodic information (e.g., a specific memory of the last encounter with that person). Finally, the target's name may be retrieved.

The same sequence of stages has also been proposed to characterize voice recognition (Belin, Fecteau, & Bédard, 2004; Ellis, Jones, & Mosdell, 1997). Although the face is a powerful cue for identifying people, the voice is also a means of identification in everyday life. Recently, researchers have started to investigate the retrieval of both semantic and episodic information following recognition of a familiar voice and, more specifically, to compare the retrieval of such information following face recognition with that following voice recognition. Most studies that have conducted such a comparison have shown that it is easier to retrieve semantic and episodic information when recognizing a familiar face than when recognizing a familiar voice.

Typically, participants in these studies are presented with a set of stimuli. Half of the stimuli are celebrities' faces or voices, whereas the other stimuli are unfamiliar faces or voices. For each stimulus, the participant's first task is to judge whether or not the stimulus is familiar (a *familiarity judgment*). If the participant's response is positive, he or she is then asked to provide semantic details about the recognized person—usually the person's occupation (i.e., a retrieval of semantic information). Finally, the participant is invited to name the target person. In almost all the experiments we report here, the stimulus domain (face vs. voice) has been a between-participants factor. In other words, different groups of participants were presented with faces and with voices (there are only two exceptions where a within-participants design was employed: Barsics & Brédart, 2012a; Hanley & Damjanovic, 2009, Experiment 2).

The Overall Recognition-Performance Issue

The comparison of information retrieval from faces and from voices started with a study by Hanley, Smith, and Hadfield (1998) showing that semantic information was less likely to be recalled by participants whose task was to recognize celebrities from their voice than by those whose task was to recognize the same famous people from their face. In other words, experiences of finding a voice familiar without retrieving any further information about the voice's owner (i.e., *familiarity-only*

Corresponding Author:
Serge Brédart, Department of Psychology: Cognition and Behavior, University of Liège, 4000 Liège, Belgium
E-mail: serge.bredart@ulg.ac.be

experiences) were much more common when a voice had been recognized than when a face had been recognized.

Unfortunately, a potential artifact complicated the interpretation of these results. Indeed, on the one hand, the overall rate of recognition was lower for voices (60%–70%) than for faces (more than 90%) in this study. On the other hand, the rate of false alarms was higher for voices (about 30%) than for faces (about 20%). This made it difficult to directly compare the amount of biographical information retrieved from recognized faces and from recognized voices. It was possible that the more frequent occurrence of familiarity-only experiences for voices did not reflect a genuine processing advantage for faces but simply reflected the fact that participants produced "familiar" responses on the basis of guesswork more often in the voice condition than in the face condition.

To avoid this problem, Hanley and Turner (2000) brought face-recognition performance down to the same level as voice-recognition performance by presenting participants with blurred faces. In such conditions, the researchers found that the rate of familiarity-only experiences was similar when blurred faces were recognized and when voices were recognized. In this case, the recall of a target's occupation was not more difficult after voice recognition than after face recognition.

At first sight, by avoiding a methodological bias, the Hanley and Turner (2000) study invalidated the view that semantic information is more accessible from faces than from voices. But this conclusion is premature, because more recent studies have cast doubt on Hanley and Turner's (2000) results, suggesting that these results are themselves very probably due to serious methodological problems.

Controlling for the Content of Speech Extracts

Eliminating nonfacial cues to identity in photographs is relatively easy and is commonly done in face-recognition research. Usually, concealing background and sartorial cues by using image-manipulation software is sufficient. Controlling for contextual cues to a target's identity in a speech extract is much more difficult. In fact, Hanley and Turner (2000) did not strictly control the content of the speech extracts used for the recognition of voices. It is therefore possible that some extracts used in their study provided contextual cues, leading to a high level of accuracy in the recall of targets' occupations in the absence of genuine person identification from the voice itself. Specifically, as noted by Hanley and Damjanovic (2009), 40% of the original celebrity-voice samples used in Hanley and Turner's (2000) study could be matched to the target's correct occupation on the basis of guesswork alone.

Several more recent studies have used a recognition procedure similar to that used by Hanley and Turner but followed the guidelines of Van Lancker, Kreiman, and Emmorey (1985) and Schweinberger, Herholz, and Steif (1997) to limit the extent to which the speech content of the extracts could give clues to the targets' identity. For instance, each speech sample had to be free of catchphrases and identifying sounds, such as the sounds of a studio audience or a theme tune. Under these circumstances, results are unambiguous. *All* the more recent studies that have strictly controlled the content of spoken extracts have unambiguously indicated that semantic information about familiar people (e.g., a familiar person's occupation) is easier to retrieve when recognizing a face than when recognizing a voice (e.g., Barsics & Brédart, 2011; Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009), even though these studies used blurred faces as stimuli to ensure that overall recognition performance was similar for both types of stimuli.

Another strategy that has been used to control the content of speech extracts is the presentation of faces and voices of personally familiar targets (e.g., participants' teachers) rather than celebrities, a method that makes it possible to have all the target persons speak the same words for the extracts. In one of our studies (Brédart, Barsics, & Hanley, 2009), these familiar targets all read the same scripted monologue (the first article of the United Nations' Universal Declaration of Human Rights) in their speech extracts. Again, results showed a memory advantage for faces over voices: Semantic information about targets (e.g., the subject taught by the target teacher) was more easily recalled after the recognition of their blurred faces than after the recognition of their voices.

The Frequency-of-Exposure Issue

A third problem that could limit the comparability of voices and faces was acknowledged in the report of the first study that compared the retrieval of information from faces and from voices (Hanley et al., 1998): We see celebrities' faces in the media more frequently than we hear their voices. We presumably see the faces of actors and actresses, politicians, and athletes without hearing their voices—in magazines, in newspapers, and even on Web sites—much more frequently than we hear these celebrities' voices without seeing their faces. Therefore, it is possible that the observed memory advantage for faces over voices is merely a consequence of the fact that we are more often exposed to famous people's faces than to their voices, and not to an intrinsically privileged access to semantic memory from the face-recognition system.

The use of faces and voices of personally familiar persons as stimuli has helped researchers to bypass this problem, because when such people are encountered, they are usually both seen and heard. We thought that the faces and voices of participants' teachers were particularly interesting stimuli (Brédart et al., 2009). Indeed, even if this is difficult to quantify, although students see their teachers, they also often hear their voices without seeing their faces when taking notes and looking at slides. Hence, the problem of the greater exposure to target faces seemed to be at least reduced when teachers' faces and voices were used as stimuli. The two studies that have used this kind of stimuli replicated the face-advantage effect: The retrieval of semantic information was substantially

better among students who recognized their teachers from their normal and blurred faces than among students who recognized their teachers from their voices (Barsics & Brédart, 2011; Brédart et al., 2009).

A more powerful way to control the frequency of exposure to faces and voices is to use an associative-learning paradigm in which participants have to associate semantic information and names with pre-experimentally unfamiliar faces or voices. In fact, the use of such learning paradigms allows researchers to strictly control both the frequency of exposure to the two types of stimuli and of the content of speech extracts. For instance, in one recent study (Barsics & Brédart, 2012b), a name and an occupation were presented in association with, respectively, a face, a voice, or both a face and a voice to three different groups of participants. Each association was repeated four times. After this learning phase, a cued-recall task started. Each learned stimulus (face, voice, or both face and voice, depending on condition) was presented, and the participants' task was to provide the occupation and the name for each. Results indicated that performance was significantly lower among participants in the voice-only condition than among participants in the face-only and face-plus-voice conditions. Therefore, the advantage of faces over voices remained even when the frequency of exposure to the two kinds of stimuli was strictly equivalent.

The Face Advantage Extends to the Retrieval of Episodic Memory

In the studies reviewed above, the central goal was to compare the retrieval of identity-specific semantic information after face recognition and after voice recognition. A couple of studies assessed whether the retrieval of episodic information about familiar people was also easier after face recognition than after voice recognition. Damjanovic and Hanley (2007) compared the extent to which the recognition of a face or a voice was accompanied by the recollection of a specific episode in which that face or voice was present (i.e., *remember* responses; e.g., "I remember watching her funeral on TV and seeing all the flowers" following the recognition of Princess Diana) or by a simple feeling of familiarity or the mere knowledge of a fact about the target that did not encompass such a recollection (i.e., *know* responses; e.g., "I know that's Diana and her sons are William and Harry, but I don't have a specific memory for her").

These studies demonstrated that both normal and blurred faces elicited more *remember* responses than *know* responses. Conversely, voices elicited more *know* responses than *remember* responses. In addition, there were more *remember* responses following the recognition of a blurred face than following the recognition of a voice, although the rates of recognition were again similar for blurred faces and voices. Using the faces and voices of personally familiar people (i.e., participants' teachers) as stimuli, we also observed that participants were more likely to retrieve specific memories related to target persons following the recognition of their blurred faces than

following the recognition of their voices (Barsics & Brédart, 2011).

Attempting to Explain the Face Advantage

We have shown that by applying adequate methodological controls, numerous studies have consistently indicated that retrieving semantic as well as episodic information about familiar persons is easier following the recognition of these persons' faces than following the recognition of these persons' voices, even when the overall recognizability of faces and voices was matched. This advantage of faces over voices is robust. Indeed, it occurs for different categories of target persons (celebrities, personally familiar people, and newly encountered persons). Moreover, it occurs regardless of whether the domain of stimuli (faces vs. voices) is a between-participants or a within-participants factor.

The most common explanation for the face advantage in semantic-information retrieval is that the associative links between the representation of a face and semantic memory are stronger than the corresponding links between the representation of a voice and semantic memory (Damjanovic, 2011; Gainotti, Barbier, & Marra, 2003; Hanley & Damjanovic, 2009). Similarly, it has been hypothesized that the connections between the face-recognition system and episodic memory are stronger than those between the voice-recognition system and episodic memory (Damjanovic & Hanley, 2007). As an alternative to this differential-strength-of-connections hypothesis, another account based on stimulus confusability (Stevenage, Hugill, & Lewis, 2012) has been put forward. It is possible that we distinguish between faces more easily than we distinguish between voices. Some earlier studies suggested that voice-discrimination skills are poor in comparison with face-discrimination skills (e.g., Yarmey, Yarmey, & Yarmey, 1994).

Considering that distinctive stimuli are less confusable than typical ones are, we compared the retrieval of semantic and episodic information from distinctive faces and voices and from typical faces and voices (Barsics & Brédart, 2012a). Although the retrieval of information was better following the recognition of distinctive stimuli than following the recognition of typical stimuli, typical faces yielded a better recall of information than distinctive voices did. Therefore, the advantage of faces over voices persisted even when distinctiveness was manipulated in favor of voices. However, it is possible that the distinctive voices in our experiment were still more difficult to discriminate than typical faces were. Further research is needed to investigate the possible role of confusability in the advantage of faces over voices.

It remains that, up to now, the most commonly invoked explanation for this face advantage is that connections between face representations and episodic- or semantic-memory systems are stronger than connections between voice representations and those memory systems. Future research will be needed to determine why such connections would be stronger for faces. In their differential-utilization account, Stevenage et al. (2012) recently suggested that one reason for the face

advantage might be that voices are experienced less often than faces. Moreover, these authors argued that when we are exposed to both the face and the voice of a person, we may extract the person's identity from the face but extract the meaning of the person's speech from the voice. Therefore, even in cases of face and voice co-occurrence, the face would be dominantly used to identify a person. In addition to this advantage of faces over voices, characterizing further how information from faces and voices interact during person recognition remains particularly important (Stevenage et al., 2012; for a review, see Campanella & Belin, 2007).

Recommended Readings

Barsics, C., & Brédart, S. (2011). (See References). A study demonstrating that the recall of episodic and semantic information about personally familiar individuals is easier following the recognition of faces (even blurred faces) than following the recognition of voices.

Bruce, V., & Young, A. (1986). (See References). The first description of the most influential model of face recognition to be developed in the past 25 years.

Gainotti, G. (2011). What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia*, *49*, 2273–2282. A comprehensive synthesis of work on person recognition from faces and voices, including neuropsychological literature on recognition disorders.

Hanley, J. R., & Damjanovic, L. (2009). (See References). A paper highlighting the risks in overlooking the semantic content of speech extracts for celebrity targets, including the potential for inadequate interpretation of data.

Hanley, J. R., Smith, T. S., & Hadfield, J. (1998). (See References). A pioneering paper that directly compared the recognition of familiar voices and familiar faces.

Acknowledgments

Declaration of Conflicting Interests

Funding

References

Barsics, C., & Brédart, S. (2011). Recalling episodic information about personally known faces and voices. *Consciousness and Cognition*, *20*, 303–308.

Barsics, C., & Brédart, S. (2012a). Access to semantic and episodic information from faces and voices: Does distinctiveness matter? *Journal of Cognitive Psychology*, *24*, 789–795

Barsics, C., & Brédart, S. (2012b). Recalling semantic information about newly learned faces and voices. *Memory*, *20*, 527–534.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135.

Brédart, S., Barsics, C., & Hanley, R. (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology*, *21*, 1013–1021.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*, 535–543.

Damjanovic, L. (2011). The face advantage in recalling episodic information: Implications for modeling human memory. *Consciousness and Cognition*, *20*, 309–311.

Damjanovic, L., & Hanley, J. R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory & Cognition*, *35*, 1205–1210.

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and intermodal priming of familiar faces and voices. *British Journal of Psychology*, *88*, 143–156.

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.

Hanley, J. R., & Damjanovic, L. (2009). It is more difficult to retrieve a familiar person's name and occupation from their voice than from their blurred face. *Memory*, *17*, 830–839.

Hanley, J. R., Smith, T. S., & 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.

Schweinberger, S. R., Herholz, A., & Steif, V. (1997). Auditory long term memory: Repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology*, *50A*, 498–517.

Stevenage, S. V., Hugill, A. R., & Lewis, H. G. (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, *24*, 409–419.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics*, *13*, 19–38.

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology*, *8*, 453–464.