

INVITED REVIEW

Person Recognition Is Easier from Faces than from Voices

Catherine Barsics*

This article reviews a number of recent studies that systematically compared the access to semantic and episodic information from faces and voices. Results have showed that semantic and episodic information is easier to retrieve from faces than from voices. This advantage of faces over voices is a robust phenomenon, which emerges whatever the kind of target persons, might they be famous, personally familiar to the participants, or newly learned. Theoretical accounts of this face advantage over voice are finally discussed.

Keywords: person recognition; face; voice; episodic memory; semantic memory

When we recognize a person, we can retrieve different kinds of information about her/him: semantic information (e.g., the person's occupation), episodic information, such as a memory of a specific occasion on which this person has previously been encountered and finally, lexical information (i.e., the person's name). During the last three decades, most research investigating person recognition has mainly investigated access to semantic and lexical information from faces, building on the Bruce and Young's (1986) seminal model of face processing.

However, other cues to person identity, such as gesture, gait, body shape, and voice, deserve to receive more attention, at least because in some cases facial information is not available while the correct identification of the person is of crucial importance.

Moreover, in everyday social interactions, the face is rarely the only cue available to identify a person. A growing number of studies now aim at complementing existing models of person recognition by characterizing the mechanisms of multi-modal information integration (for a recent review, see Schwartz, 2014). In this line, some models of person recognition, endeavoring to encompass the fundamentally multi-modal nature of person recognition, have integrated a voice recognition pathway in their architecture (Belin, Fecteau, & Bédard, 2004; Campanella & Belin, 2007; Ellis, Jones, & Mosdell, 1997).

In the present paper, we review a number of studies that systematically compared the access to semantic and episodic information following the recognition of faces and voices. These studies have faced distinct methodological difficulties inherent to the necessity of establishing appropriate ways of comparing voice and face recognition, and of adapting to the specific use of voice stimuli (see

* Swiss Center for Affective Sciences, University of Geneva-CISA, Geneva, Switzerland
catherine.barsics@unige.ch

also Brédart & Barsics, 2012; Damjanovic, 2011). The main finding of these studies is the fact that the retrieval of semantic and episodic information is easier from faces than from voices.

Comparing famous faces and voices recognition

In 1998, in a pioneering paper, Hanley, Smith, and Hadfield directly assessed the relative access to semantic information from familiar faces and from familiar voices. They presented participants with either famous faces or famous voices, and assessed their level of recognition for these items, along with their ability to retrieve the occupations of the celebrities. Results showed that participants were far better at recognizing celebrities from their faces than from their voices. Indeed, 70% of the famous voices were identified as familiar in the voice condition, but 94% in the face condition (Experiment 1). Moreover, in the voice condition, participants recalled significantly less occupations than in the face condition: 63% of the voices deemed familiar were accompanied with the correct recall of the occupation in the voice condition, compared with 92% in the face condition. In other terms, participants experienced more “familiarity-only” feelings in response to voices than to faces. Familiarity-only feelings are characterized by a sense of familiarity for the target despite of an inability to access any related semantic information.

However, this pattern of results was difficult to interpret. Indeed, there were significantly more false alarms (i.e. misclassification of a non-famous person as famous) in the voice than in the face condition. Discrimination was significantly better in the face than in the voice condition. Furthermore, decision criterion was significantly higher in the voice than in the face condition. Consequently, participants might have judged a high number of items in the voice condition as familiar on the basis of mere guesswork. It would therefore be not surprising they would not

be able to retrieve names and occupations associated with these items.

In order to circumvent the difficulty raised by the higher level of overall performance for faces than for voices, Hanley and Turner (2000) attempted to match the overall recognition rates in the face and voice conditions by reducing the level of performance for faces. To that purpose, the authors resorted to an original procedure. They presented blurred faces to participants, with a level of blur calibrated such as the familiarity for blurred faces equated that for voices. They assessed whether, in such conditions, it would still be easier to recall occupations from familiar faces than voices. Recall of occupations was significantly higher in the standard face than in the blurred face and voice conditions. However, in these last two conditions, in which level of familiarity was equated, there was no difference between the proportion of occupations and names recalled when a face or voice was recognized at each of the levels of familiarity. The addition of a blurred face condition created particularly ideal conditions in order to compare faces with voices: the number of hits and false alarms were similar in the blurred face and voice condition, which were also matched for both sensitivity and criterion. These results casted doubts on the face advantage over voice that had previously been highlighted (Hanley, Smith, & Hadfield, 1998).

So far, the investigation of familiar face recognition processes was focused on the retrieval of lexical and semantic information associated with recognition (e.g. Hanley & Cowell, 1988 ; Hanley et al., 1998 ; Hanley & Turner, 2000; Hay et al., 1991; Young et al., 1985), rather than on the retrieval of episodic information. Before a study by Damjanovic and Hanley, in 2007, no data were available regarding the retrieval of episodic information associated with person recognition. This question gained in importance in the light of Westmacott and Moscovitch's (2003) and Westmacott, Black, Freedman, and Moscovitch's (2003)'s findings. Their studies

indicated that episodic memory might play a more significant role in person identification than previously assumed. In this perspective, Damjanovic and Hanley (2007) compared not only the retrieval of semantic and lexical information (i.e. identity-specific details and name, respectively), but also the retrieval of episodic information, following the presentation of famous faces and voices. In the manner of Westmacott and Moscovitch (2003), Damjanovic and Hanley (2007) implemented the Remember/Know paradigm (Tulving, 1985; see also Gardiner, Ramponi, & Richardson-Klavehn, 1998) to the recognition of famous faces and voices. Although used in experiments dedicated to face recognition study (e.g. Konstantinou & Gardiner, 2005), this procedure had never been employed to explore the recognition of pre-experimentally familiar faces and voices.

The Remember/Know paradigm allowed Damjanovic and Hanley (2007) to compare the states of awareness associated with the recognition of famous standard faces, blurred faces and voices. They examined to what extent the recognition of a familiar face or voice was accompanied by the recollection of a specific episode implying the target person ("Remember" responses) or by a mere familiarity judgment devoid of any recollection ("Know" responses). Episodic information (i.e. Remember responses) was significantly better retrieved in the standard face than in the blurred face condition, and in the blurred face than in the voice condition. In other terms, at similar level of familiarity and overall recognition, it is still easier to recall episodic information from blurred faces than from voices. The probability that a famous standard face that had been found familiar would elicit episodic information was 55%. Corresponding figures were 46% in the blurred face condition and 31% in the voice condition. On every occasion participants made an R response, they were also able to retrieve semantic information about the target person. Semantic information was significantly more likely to be recalled

from blurred faces than voices. This observation contrasts with the results of Hanley and Turner (2000) but is consistent with the earlier results of Hanley, Smith, and Hadfield (1998). Damjanovic and Hanley (2007) argue that their experiment counted a number of methodological improvements compared with the Hanley and Turner (2000) study. Mainly, the material was selected more carefully, and they do not exclude that it might be possible that Hanley and Turner's (2000) voice material might have carried along contextual or semantic cues to the person identity, plausibly leading to an artificially high level of correct occupations recalled in the voice condition on such a basis.

Facing methodological difficulties

Actually, the material created by Hanley and Turner (2000) included some extracts in which celebrities were playing their own role, whereas the stimuli of the Damjanovic and Hanley (2007) study came from lifestyle documentary programs. In order to ensure that the speech content pronounced by the target persons was devoid of any clue to the person identity, Damjanovic and Hanley's (2007) stimuli had been prepared following the guidelines from Van Lancker, Kreiman, and Emmorey (1985) and Schweinberger, Herholz, and Steif (1997). Each speech sequence was thus free of catchphrases, background noises (e.g. studio audience or jingle), and was pronounced in an emotionally neutral tone and with normal intensity. A pilot study proved that it was not possible to guess the identity of the speakers from the content of their speech, which had been presented to an independent group of participants in a written form. In order to explain the discrepancies between the two studies, Hanley and Damjanovic (2009) used Damjanovic and Hanley's (2007) material while providing participants with instructions similar to those of Hanley and Turner (2000). In this context, significantly more occupations and names were retrieved in the blurred face condition than in the voice

condition, although the recognition performance was analogous in both conditions. Hence, these results differed from those obtained by Hanley and Turner (2000) but they were however highly consistent with those obtained by Damjanovic and Hanley (2007), despite the fact that participants did not have, this time, to recall any episodic information.

Therefore, the contrasted results obtained in previous studies investigating the recall of semantic and episodic information from celebrities' faces and voices might have been a consequence of methodological difficulties (Damjanovic & Hanley, 2007; Hanley et al., 1998; Hanley & Damjanovic, 2009; Hanley & Turner, 2000). In face-recognition research, removing non-facial cues, such as background and sartorial items, in face photographs is relatively easy and is usually carried out by resorting to image editing softwares. Other kinds of precautions must be applied when speech extracts have to be used in person recognition paradigms.

Voice material selection has to follow careful procedures. Guidelines from Van Lancker, Kreiman, and Emmorey (1985) and Schweinberger, Herholz, and Steif (1997) are extremely useful in this regard. Selected speech sequences have to be free of catch-phrases or background noises, pronounced in an emotionally neutral tone and with normal intensity. Pilot studies, aimed at ensuring that celebrities cannot be recognized from the speech content of the voice stimuli, highly contribute to guarantee the adequacy of the voice material. The methodology used to select adequate voice material improved across studies, as clearly demonstrated by Hanley and Damjanovic (2009), who showed that 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.

Nevertheless, at that point, it has been suspected that famous people might perhaps not constitute the most ideal stimuli in

order to investigate the retrieval of semantic information from faces and voices. Indeed, as already acknowledged by Hanley et al. (1998), we are certainly more frequently exposed to celebrities' faces than to their voices. Most of the time, in the media, we see faces of actors, athletes or politicians without hearing their voices. The observed advantage of faces over voices might therefore result from our privileged frequency of exposure to famous faces compared with famous voices.

Investigating personally familiar faces and voices recognition

The matter of matching faces and voices in terms of familiarity and overall recognition has been resolved by blurring the faces to equate overall performance for faces and voices, as first introduced by Hanley and Turner (2000). In order to bypass the differential frequency of exposure to faces and voices issue, faces and voices of *personally familiar* people have been considered to constitute interesting stimuli to assess the access that they provide to semantic and episodic information. Indeed, when we meet personally familiar people, we are usually exposed to both their faces and voices. Moreover, using stimuli that are personally familiar to the participants also allows a strict control of the voice material. The content of the spoken extracts presented to the participants can be devoid of any cue to the speakers' identities, the tone can be kept emotionally neutral and the intensity can be easily kept similar.

In this line, Brédart, Barsics, and Hanley (2009) presented students with their teachers' voices and faces, among unknown voices and faces. All speech extracts were identical (the first article of the United Nations' Universal Declaration of Human Rights). In order to equate the familiarity between faces and voices, blurred faces were also displayed in a third condition. Participants had to perform a yes/no recognition task for these three kinds of stimuli. In case of recognition, they were requested to state the target person's name and identity-specific details

such as the subject of a professor's course. Semantic information and names were better retrieved from faces than from voices, even when familiarity for faces and voices was rendered similar to that of voices by blurring the faces.

The retrieval of episodic information from personally familiar voices compared to personally familiar faces has also been investigated with such kinds of stimuli (Barsics & Brédart, 2011), in a paradigm similar to that one used by Damjanovic and Hanley (2007). When participants made a positive recognition decision to the stimuli, they were asked to specify whether they were able to retrieve any episodic information associated with the deemed familiar person, any identity-specific semantics about this person, or if they were in a familiarity-only state, or merely guessing. Again, results showed a memory advantage for faces over voices: both episodic and semantic information was more likely to be retrieved following familiar face than following familiar voice recognition. The advantage of faces over voices regarding the retrieval of episodic and semantic information remained stable even when face recognition was rendered similar to that for voices by blurring the faces.

A further way of ensuring both a strict control of the frequency of exposure to the face and voice stimuli and the absence of identity clues in the voice extracts is to resort to the use of an associative-learning paradigm. In a recent study (Barsics & Brédart, 2012a), participants had to associate pre-experimentally unfamiliar faces or voices with semantic information (i.e. occupations) and names. More precisely, during the learning phase, names and occupations were presented to the participants along with either a face, a voice, or both a face and a voice, depending on the condition participants were assigned in. A cued-recall task followed, in which participants were requested to provide the name and occupation in response to the presentation of the associated face, voice or both face and voice. In such a procedure, the frequency

of exposure to faces and voices was strictly equivalent, but the face advantage over voice nevertheless emerged, as performance was significantly lower in the voice-only condition than in the face-only and face-plus-voice conditions. In addition, neither benefit nor disadvantage emerged from the concomitant presentation of faces and voices.

Explaining the face advantage

All the studies that have applied appropriate methodological controls have consistently shown a face advantage over voice, both in terms of a better access to *semantic* information from faces than from voices (Brédart et al., 2009; Hanley & Damjanovic, 2009; Hanley et al., 1998), and in terms of a better access to *episodic* information from faces than from voices (Barsics & Brédart, 2011; Damjanovic & Hanley, 2007). This face advantage has been demonstrated with famous faces and voices (Barsics & Brédart, 2012b; Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009; Hanley et al., 1998), with personally familiar faces and voices (Brédart et al., 2009; Barsics & Brédart, 2011), as well as with newly learned faces and voices (Barsics & Brédart, 2012a). In addition, this robust phenomenon occurs regardless of whether the domain of stimuli (faces vs. voices) is a between-participants or a within-participants factor, and it remains stable even when the overall recognizability of faces and voices is pre-experimentally equated.

The face advantage over voice is compatible with both Bruce and Young's (1986) model and with Interactive Activation and Competition (IAC) models of person recognition (e.g. Brédart, Valentine, Calder, & Gassi, 1995; Burton, Bruce, & Johnston, 1990; Stevenage, Hugill, & Lewis, 2012) - such a reasoning has been advocated before (e.g. Hanley et al., 1998). Here, the critical difference between these two models lies in the locus of the familiarity decision, which is located in a more distal point of the framework in IAC models than in the Bruce and Young's (1986) model (for recent reviews

comparing these accounts, see Gainotti, 2011; Hanley, 2014). Within the seminal person recognition model of Bruce and Young (1986), familiarity decisions are thought to occur at the level of Face Recognition Units (FRUs) and Voice Recognition Units (VRUs). Hence, a familiarity-only response would appear when a FRU or a VRU has been sufficiently activated, but without any activation of the related Person Identity Node (PIN). If connections between VRUs and PINs are weaker than between FRUs and PINs, the Bruce and Young (1986) model can account for the fact that semantic information is more likely to be retrieved from familiar faces than voices. In IAC models (e.g. Brédart, Valentine, Calder, & Gassi, 1995; Burton, Bruce, & Johnston, 1990; Stevenage, Hugill, & Lewis, 2012) familiarity decisions are assumed to emerge at the level of the PINs. Since PINs are modality-free levels, the retrieval of semantic information from faces and voices should be similar from familiar faces and voices, especially when they are matched for familiarity. The face advantage and IAC models can nevertheless be conciliated if, again, one postulates that the connections between VRUs and PINs are weaker than between FRUs and PINs. When a VRU passes on its activation to its associated PIN, the received input might be sufficient to trigger a positive familiarity decision, but not strong enough to activate the related SIUs, therefore giving rise to a familiarity-only feeling. This kind of interpretation has been previously suggested in order to account for a neuropsychological case, in a study describing a patient who displayed developmental impairment of both access to biographical information from faces and face naming (Van der Linden, Brédart, & Schweich, 1995).

Alternatively to the account of the face advantage in terms of stronger connections between the representation of a face and semantic memory than between the representation of a voice and semantic memory, Stevenage, Hugill, and Lewis (2012) provided

an interesting account in order to explain why the face advantage occurs. It encompasses two potential mechanisms by which the discrepancy between faces and voices might emerge.

First, Stevenage, Hugill, and Lewis (2012) suggest that voice links with PINs might be weaker because of the *differential utilization* of faces and voices. Indeed, as previously acknowledged by Hanley et al. (1998) and by Hanley and Turner (2000), in everyday life, we are much more exposed to famous faces than to famous voices through media. Consequently, the connections between VRUs and PINs might benefit from less time and opportunity to strengthen through repeated exposures, in comparison with FRUs and PINs. This was the reason why we resorted to the use personally familiar stimuli in order to compare the access to semantic and episodic information from faces and voices (Brédart et al., 2009; Barsics & Brédart, 2011). When we meet personally familiar people, we usually both see and hear them. In the particular case of teachers presented to their students during recognition tests, one might even think that they have been massively exposed to their voices, as when taking notes during lectures, students would listen more to their teachers' voices than look at them. However, even in this context, the face advantage remains, regarding the retrieval of semantic as well as of episodic information.

The second argument of the *differential utilization* account (Stevenage et al., 2012) posits that we might spontaneously rely on the face in order to derive person identity, whereas we would actually use the voice as a means to extract non-identity based information. Intuitively indeed, it seems that when we hear a voice, we tend to be more concerned with the content of the speech than with its auditory characteristics. While these two types of processing are of course not mutually exclusive, one should be aware of the possibility that even when face and voice are both available, we might tend to rely more on face than on other cues in

order to extract the person identity, which in turn would lead to the superiority of face over voice when these cues are presented separately for recognition. This account is highly compatible with Le Breton's (2003) observation that in our society, individuation is strongly related to identifying people by their faces and names. Furthermore, Stevenage, Howland, and Tippelt (2011) showed that faces accompanying voices at study interfered with subsequent voice identification, but voices accompanying faces at study did not interfere with subsequent face identification. In the same vein, Cook and Wilding's (1997) results indicated that the presence of a face at study impaired subsequent voice recognition, both when the face was again present at test as when it was absent. Such an interference effect had also been demonstrated by McAllister, Dale, Bregman, McCabe and Cotton (1993), whose results showed that voice recognition was more impaired by face co-presentation than was face recognition by voice co-presentation. Finally, results such as those of Barsics and Brédart (2012a), described hereinabove, are also highly compatible with the idea that faces would be preferentially used for the purpose of identification, whereas voices would be used to extract and process non-identity based information.

The second account of the weaker voice pathway proposed by Stevenage et al. (2012) lies in the potential *differential confusability* that we might experience towards faces and voices. The differential confusability refers to the fact that voices are characterized by more perceptual confusability than faces. In line with this idea, Barsics and Brédart (2012b) used the relative distinctiveness of faces and voices as a means to manipulate confusability. They assessed the retrieval of semantic and episodic information from distinctive faces and voices and from typical faces and voices, with familiarity being matched across domains of stimuli. Results showed that distinctiveness is indeed an important factor, as semantic information

was better retrieved from distinctive than typical stimuli. However, distinctiveness had less impact than domain on the recall of semantic details, since more semantic information was retrieved from typical faces than from distinctive voices. Thus, the face advantage persisted even when distinctiveness was manipulated in favor of voices. Nonetheless, it remains possible that the distinctive voices in this experiment were still more confusable than typical faces, and further research should allow to better characterize the potential role of confusability in the advantage of faces over voices.

From our standpoint, the advantage of faces over voices regarding the access to semantic and episodic information could be contingent on our degree of expertise with these two domains of cues to the person identity. More precisely, we posit that the discrepancy between face and voice recognition, and the subsequent likelihood to retrieve episodic and semantic information from them, could be attributed to our *differential expertise* for these two kinds of stimuli. This standpoint has the advantage of merging the two Stevenage et al.'s (2012) accounts in only one. Indeed, the higher frequency of exposure to faces than to voices (i.e. differential utilization) should influence our level of expertise for these two domains of stimuli. This differential expertise would therefore give rise to distinct abilities of discrimination of faces and voices, which are directly related to their differential confusability. Consequently, the relative expertise towards faces and voices could be considered as a potential explanation for their distinct proneness to be recognized and to yield the retrieval of semantic and episodic information. Although a great deal of research has focused on visual expertise for faces (e.g. Gauthier, Skudlarski, Gore, & Anderson, 2000), only a few studies have investigated the field of auditory expertise, especially when it comes to voice recognition (e.g. Chartrand, Peretz, & Belin, 2008). Kreiman, Geratt, Precoda, and Berke (1992) showed

that naive listeners relied on the fundamental frequency of phonation to make similarity judgments on voices, whereas experts also used formants (see also Baumann & Belin, 2010). Thus, we think that the investigation of voice expertise is rather appealing, both as a mean to explore the face advantage over voice and in order to understand the fundamental processes underlying voice recognition and discrimination.

Conclusion

Both faces and voices convey speech, affective, identity, and otherwise socially relevant information. Given these shared properties of faces and voices, they have been analogized as “auditory faces” (e.g. Belin, Bestelmeyer, Latinus, & Watson, 2011). In favor of this assertion, Belin et al. (2011) reviewed a number of evidence supporting similar and interacting functional architecture for the cerebral processing of faces and voices. Notwithstanding, the consistent finding that voices are much more difficult to recognize than faces, is not in favor of a conceptualization of voices as “auditory faces”. Indeed, numerous studies have showed that faces are better recognized than voices, and that the face advantage over voice, regarding the retrieval of both semantic and episodic information, is a robust phenomenon. As suggested by Stevenage et al. (2012), faces could be preferentially processed in relation to identity, while voices could be preferentially processed in respect to the extraction of non-identity based information, such as the speech content or the emotion. This would allow us to achieve multiple processing by relying on all available cues in a non-redundant manner. In other terms, the so-called ‘face advantage’, i.e., a preferential process of faces as a cue to extract the person’s identity and to access related information stored in memory, might actually favor a ‘voice advantage’, that would consist in the simultaneous process of voices in order to extract subtly nuanced emotions and the meaning of the speech. Thereby, we argue that the fact that voices represent less salient cues to the identity than faces, could

be considered as an asset and not as a malfunction of our cognitive system.

Acknowledgments

I would like to thank Serge Brédart, Salvatore Campanella and Gilles Pourtois for organizing the Symposium “Person perception from face and voice”.

References

- Barsics, C., & Brédart, S.** (2011). Recalling episodic information about personally known faces and voices. *Consciousness and Cognition, 20*(2), 303–308. DOI: <http://dx.doi.org/10.1016/j.concog.2010.03.008>
- Barsics, C., & Brédart, S.** (2012a). Recalling semantic information about newly learned faces and voices. *Memory, 20*(5), 527–534.
- Barsics, C., & Brédart, S.** (2012b). Access to semantic and episodic information from faces and voices: Does distinctiveness matter? *Journal of Cognitive Psychology, 24*(7), 789–795.
- Baumann, O., & Belin, P.** (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research, 74*(1), 110–120. DOI: <http://dx.doi.org/10.1007/s00426-008-0185-z>
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R.** (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711–725. DOI: <http://dx.doi.org/10.1111/j.2044-8295.2011.02041.x>
- Belin, P., Fecteau, S., & Bedard, C.** (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135. DOI: <http://dx.doi.org/10.1016/j.tics.2004.01.008>
- Brédart, S., & Barsics, C.** (2012). Recalling semantic and episodic information from faces and voices: A face advantage. *Current Directions in Psychological Science, 21*(6), 378–381. DOI: <http://dx.doi.org/10.1177/0963721412454876>

- Brédart, S., Barsics, C., & Hanley, R.** (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology, 21*(7), 1013–1021. DOI: <http://dx.doi.org/10.1080/09541440802591821>
- Brédart, S., Valentine, T., Calder, A., & Gassi, L.** (1995). An interactive activation model of face naming. *Quarterly Journal of Experimental Psychology, 48A*(2), 466–486. DOI: <http://dx.doi.org/10.1080/14640749508401400>
- Bruce, V., & Young, A.** (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305–327. DOI: <http://dx.doi.org/10.1111/j.2044-8295.1986.tb02199.x>
- Burton, A. M., Bruce, V., & Johnston, R. A.** (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*(3), 361–380. DOI: <http://dx.doi.org/10.1111/j.2044-8295.1990.tb02367.x>
- Campanella, S., & Belin, P.** (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*(12), 535–543. DOI: <http://dx.doi.org/10.1016/j.tics.2007.10.001>
- Chartrand, J.-P., Peretz, I., & Belin, P.** (2008). Auditory recognition expertise and domain specificity. *Brain Research, 1220*, 191–198. DOI: <http://dx.doi.org/10.1016/j.brainres.2008.01.014>
- Cook, S., & Wilding, J.** (1997). Earwitness Testimony 2: Voices, faces, and context. *Applied Cognitive Psychology, 11*(6), 527–541. DOI: [http://dx.doi.org/10.1002/\(SICI\)1099-0720\(199712\)11:6<527::AID-ACP483>3.0.CO;2-B](http://dx.doi.org/10.1002/(SICI)1099-0720(199712)11:6<527::AID-ACP483>3.0.CO;2-B)
- Damjanovic, L.** (2011). The face advantage in recalling episodic information: Implications for modeling human memory. *Consciousness and Cognition, 20*(2), 309–311. DOI: <http://dx.doi.org/10.1016/j.concog.2010.06.002>
- Damjanovic, L., & Hanley, J. R.** (2007). Recalling episodic and semantic information about famous faces and voices. *Memory and Cognition, 35*(6), 1205–1210. DOI: <http://dx.doi.org/10.3758/BF03193594>
- 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*(1), 143–156. DOI: <http://dx.doi.org/10.1111/j.2044-8295.1997.tb02625.x>
- 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*(9), 2273–2282. DOI: <http://dx.doi.org/10.1016/j.neuropsychologia.2011.04.027>
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A.** (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7*(1), 1–26. DOI: <http://dx.doi.org/10.1006/ccog.1997.0321>
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W.** (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197. DOI: <http://dx.doi.org/10.1038/72140>
- Hanley, J. R.** (2014). Accessing stored knowledge of people from faces, names and voices: A review. *Frontiers in Bioscience, 198*–207. DOI: <http://dx.doi.org/10.2741/E702>
- Hanley, J. R., & Cowell, E. S.** (1988). The effects of different types of retrieval cues on the recall of names of famous faces. *Memory and Cognition, 16*(6), 545–555. DOI: <http://dx.doi.org/10.3758/BF03197056>
- 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*(8), 830–839. DOI: <http://dx.doi.org/10.1080/09658210903264175>
- 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(5), 179–195. DOI: <http://dx.doi.org/10.1080/713755751>
- 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(4), 1105–1116. DOI: <http://dx.doi.org/10.1080/02724980050156317>
- Hay, D. C., Young, A. W., & Ellis, A. W.** (1991). Routes through the face recognition system. *Quarterly Journal of Experimental Psychology*, 43A(4), 761–791. DOI: <http://dx.doi.org/10.1080/14640749108400957>
- Konstantinou, I., & Gardiner, J. M.** (2005). Conscious control and memory awareness when recognizing famous faces. *Memory*, 13(5), 449–457. DOI: <http://dx.doi.org/10.1080/09658210444000016>
- Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S.** (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research*, 10(35), 512–520.
- Le Breton, D.** (2003). *Des visages: Essai d'anthropologie*. Paris: Metaille.
- McAllister, H. A., Dale, R. H. I., Bregman, N. J., McCabe, A., & Cotton, C. R.** (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identification. *Basic and Applied Social Psychology*, 14(2), 161–170. DOI: http://dx.doi.org/10.1207/s15324834basp1402_3
- Schwartz, B. L.** (2014). Memory for people: integration of face, voice, name, and biographical information. In T. Perfect and S. Lindsay (Eds). *Handbook of Applied Memory* (pp. 3–19). London: Sage Publications. DOI: <http://dx.doi.org/10.4135/9781446294703.n1>
- Schweinberger, S. R., Herholz, A., & Steif, V.** (1997). Auditory long-term memory: Repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology*, 50A(3), 498–517. DOI: <http://dx.doi.org/10.1080/713755724>
- Stevenson, S. V., Howland, A., & Tippelt, A.** (2011). Interference in eyewitness and earwitness recognition. *Applied Cognitive Psychology*, 25(1), 112–118. DOI: <http://dx.doi.org/10.1002/acp.1649>
- Stevenson, S. V., Hugill, A., & Lewis, H. G.** (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, 24(4), 409–419. DOI: <http://dx.doi.org/10.1080/20445911.2011.642859>
- Tulving, E.** (1985). Memory and consciousness. *Canadian Psychology*, 26(1), 1–12. DOI: <http://dx.doi.org/10.1037/h0080017>
- Van der Linden, M., Brédart, S., & Schweich, M.** (1995). Developmental disturbance of access to biographical information and people's names: A single-case study. *Journal of the International Neuropsychological Society*, 1(6), 589–595. DOI: <http://dx.doi.org/10.1017/S1355617700000734>
- Van Lancker, D., Kreiman, J., & Emmorey, K.** (1985). Familiar voice recognition: Patterns and parameters. I: Recognition of backward voices. *Journal of Phonetics*, 13(1), 19–38.
- Westmacott, R., Black, S. E., Freedman, M., & Moscovitch, M.** (2003). The contribution of autobiographical significance to semantic memory: Evidence from Alzheimer's disease, semantic, and amnesia. *Neuropsychologia*, 42(1), 25–48. DOI: [http://dx.doi.org/10.1016/S0028-3932\(03\)00147-7](http://dx.doi.org/10.1016/S0028-3932(03)00147-7)
- Westmacott, R., & Moscovitch, M.** (2003). The contribution of autobiographical significance to semantic memory. *Memory and Cognition*, 31(5), 761–774. DOI: <http://dx.doi.org/10.3758/BF03196114>
- Young, A. W., Hay, D. C., & Ellis, A. W.** (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology*, 76(4), 495–523. DOI: <http://dx.doi.org/10.1111/j.2044-8295.1985.tb01972.x>

How to cite this article: Barsics, C. (2014). Person Recognition Is Easier from Faces than from Voices. *Psychologica Belgica*, 54(3), 244-254, DOI: <http://dx.doi.org/10.5334/pb.ap>

Submitted: 15 March 2014 **Accepted:** 28 April 2014 **Published:** 20 May 2014

Copyright: © 2014 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.

]u[*Psychologica Belgica* is a peer-reviewed open access journal
published by Ubiquity Press

OPEN ACCESS 