Masking the Identities of Celebrities and Personally Familiar Individuals: Effects on Visual and Auditory Recognition Performance

Alana C. Krix, Melanie Sauerland and Maartje J. Schreuder
Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

Abstract
The current study compared the effectiveness of masking on recognition performance for faces and voices of celebrities and personally familiar individuals. On the basis of the theory suggesting stronger memory representations for personally familiar individuals, we expected masking to be more effective for celebrities than for personally familiar stimulus persons. Furthermore, we sought to replicate the face recognition advantage with masked stimuli. Face pixelation and electronic changes of the voice pitch were applied as masking techniques, using four masking levels for each stimulus. Thirty-one undergraduate students were presented with the masked faces and voices of 10 celebrities and 10 personally familiar fellow students. As expected, more correct recognitions occurred for faces than for voices, suggesting that masking does not counteract the mechanisms causing the face recognition advantage. Unexpectedly, masking effectiveness did not differ between celebrities and personally familiar individuals. This may be due to the type of personally familiar individuals used. Within personally familiar stimuli, increased familiarity did not predict the effectiveness of masking. Whereas the highest masking level eliminated speaker recognition, masking did not fully eliminate face recognition. From a practical perspective, the findings especially question the suitability of pixelation as a means for identity concealment.

Keywords
forensic science, masking, face recognition, speaker recognition, personal familiarity, electronic voice disguise, pixelation

Corresponding author:
Alana C. Krix, Department of Clinical Psychological Science, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
Email: Alana.Krix@maastrichtuniversity.nl
In some situations, it may be essential to conceal a person’s identity. Such situations could include when a whistle-blower publishes information on illegal activities of an organization and wants to avoid penalties or when a key witness has been promised protection for providing incriminating testimony and fears retaliation. To this end, masking techniques are applied, as can often be seen on TV and in newspaper reports. These techniques can take many forms, such as pixelating or blurring the face area (e.g., Lander, Bruce, & Hill, 2001). In the case of speakers, adoption of a dialect or foreign accent, phonary adjustments (e.g., whispering), or electronic alteration of acoustic features, such as pitch (Clark & Foulkes, 2007; Künzel, 2000), can be used.

Historically, Harmon and Julesz (1973) were the first to report that pixelation of images impairs recognition. Since then, the effects of different masking techniques have been addressed in a number of studies, showing that recognition performance of both faces and voices degrades with increasing masking level, although not necessarily in a linear manner (Bindemann, Attard, Leach, & Johnston, 2013; Clark & Foulkes, 2007; Costen, Parker, & Craw, 1994, 1996; Reich & Duke, 1979). Increasing the masked area of a picture further deteriorates recognition (Demanet, Dhont, Notebaert, Pattyn, & Vandierendonck, 2007), but movement of a stimulus can counteract the detrimental effects of masking (Lander et al., 2001). Other researchers have investigated thresholds at which recognition abruptly decreases (Bachmann, 1991). For example, a resolution of eight horizontal pixels has been suggested as a lower limit for face recognition (Bhatia, Lakshminarayanan, Samal, & Welland, 1995; Bindemann et al., 2013). For electronic voice disguise, the results of one study suggest that lowering the pitch (i.e., the fundamental frequency) by eight semitones may lead to performance at chance level (Clark & Foulkes, 2007). However, this is not a unanimous finding. In two experiments, applying pitch scaling and vocal tract length scaling as masking techniques (each technique separately and in combination), Huckvale and Kristiansen (2012) had participants identify five familiar speakers. In contrast to Clark and Foulkes, a reduction of the fundamental frequency by eight semitones still generated performance above chance level. Only a rather extreme disguise, a combination of two masking techniques (increase in pitch by 12 semitones and 80% vocal tract length scaling), led to chance performance.

Despite the reduction in recognition performance caused by masking, many researchers remain critical toward its effectiveness. This is because, even at relatively high masking levels, fairly high recognition rates can be obtained, calling into question the protective value for the disguised person (Demanet et al., 2007; Huckvale & Kristiansen, 2012; Lander et al., 2001). For example, Lander et al. (2001) examined the effects of pixelation using 10 versus 20 pixels per face for different viewing distances and for still versus moving clips on recognition performance of celebrities’ faces. They observed an overall hit rate of 65%. Even in the condition with the lowest performance, the hit rate was as high as 46% and far from floor. Similarly, in a study on voice masking, Clark and Foulkes (2007) applied four different masking levels on four different speakers whom participants had been trained to identify in an earlier session. Only one of the four masking levels (−8 semitones) resulted in chance performance.

One important aspect that has not been addressed in masking research is the use of personally familiar stimuli. So far, facial masking studies utilized either celebrities’ faces (Demanet et al., 2007; Lander et al., 2001) or visually familiar faces (i.e., originally unfamiliar faces studied in an initial training session; e.g., Bachmann, 1991), but no study on facial masking effects has employed personally familiar faces of friends, family members, or colleagues. Moreover, irrespective of the modality (face or voice), famous and personally familiar stimuli have not been compared in masking studies (cf. Natu & O’Toole, 2011).
The lack of research on the effects of masking personally familiar stimuli is surprising, because the disguised persons may often be personally familiar to those from whom the identity needs to be kept secret. From face recognition research, it is known that personally familiar individuals are easily recognized (Burton, Wilson, Cowan, & Bruce, 1999), even under suboptimal viewing conditions (i.e., low-quality CCTV footage). Similarly, speaker identification performance can be high when the disguised voices come from participants’ friends (Huckvale & Kristiansen, 2012). Additionally, speaker recognition is enhanced when one engaged in a conversation with the target person before (Hammersley & Read, 1985), as is common for personally familiar, but not famous persons. Hammersley and Read suggest that in an active conversation, more information about the voice is attended to and subsequently encoded, leading to active voice memory that facilitates recognition. Face processing research has yielded evidence that there are differences in the representation of famous and personally familiar faces (Taylor et al., 2009; Walton & Hills, 2012). Personally familiar faces have been found to evoke higher levels of activation in brain areas associated with person knowledge (including personality, attitudes, and episodic knowledge) than famous faces (Gobbini, Leibenluft, Santiago, & Haxby, 2004). This means that the representations of personally familiar individuals are wider (i.e., involving more brain regions) than representations of famous faces and extend beyond visual memory. This research has led to the proposition of a face recognition model, in which recognition of a familiar person entails recognition of the visual appearance, as well as the spontaneous activation of person knowledge and an emotional response (see Gobbini & Haxby, 2007, for a review). Furthermore, Herzmann, Schweinberger, Sommer, and Jentzsch (2004) observed that face unit activation was higher for personally familiar than for famous faces, presumably because the former are more frequently encountered (but see Kloth et al., 2006). Due to the stronger and wider representation of personally familiar than that of famous individuals, masking may be less effective for the former than for the latter (cf. Lander et al., 2001).

Previous masking studies have utilized either voices or faces, but not both, within one study (e.g., Demanet et al., 2007; Huckvale & Kristiansen, 2012; Lander et al., 2001; Reich & Duke, 1979). It is a well-established finding that people are better at recognizing faces than speakers (Hanley, Smith, & Hadfield, 1998; see Barsics, 2014; Yarmey, 2007 for reviews), which is reflected in a lower diagnosticity for speaker compared with face identification procedures (Yarmey, Yarmey, & Yarmey, 1994). Several theoretical explanations have been proposed for this finding. According to one account, interference effects could be responsible such that the encoding of visual information (i.e., the face) inhibits the encoding of auditory information (i.e., the voice; McAllister, Dale, Bregman, McCabe, & Cotton, 1993). According to another account, associations of episodic memory (e.g., the name of the person) are weaker to the voice than to the face recognition system (Brédart, Barsics, & Hanley, 2009). It has also been suggested that the face recognition advantage stems from a higher expertise in using cues from the visual relative to the auditory domain to extract person information (Barsics, 2014). All accounts imply that the face recognition advantage would prevail when stimuli are masked.

It was the aim of the present study to test the effects of masking on recognition performance of famous and personally familiar persons’ faces and voices. As masking techniques, we used pixelation of the faces and, to increase the scarce database in this field (Clark & Foulkes, 2007; Huckvale & Kristiansen, 2012), electronic change of the voice pitch. Participants (university students) were asked to identify the masked faces and voices of 10 celebrities and of 10 fellow students. For this purpose, four different masking levels were applied to each stimulus, with each stimulus first being presented on the highest masking level. If the stimulus was not recognized at this level, the next lower masking level of
the stimulus was presented in turn. We were also able to collect the number of shared tutorial
groups of the participants and the students serving as stimuli, which may be considered a
proxy for familiarity. Necessitating active communication among the students, the tutorial
groups ensure frequent exposure to both faces and voices. Inclusion of this measure allowed
for a fine-grained analysis of the effects of familiarity on recognition performance of masked
stimuli.

In line with previous research (e.g., Huckvale & Kristiansen, 2012; Lander et al., 2001), we
expected that masking would not fully eliminate recognition. Drawing on face and voice
processing theory and research (Gobbini & Haxby, 2007; Hammersley & Read, 1985;
Herzmann et al., 2004), we had the following expectations for the comparison of
personally familiar and famous stimuli. The proportion of correct recognitions should be
higher and the proportion of false recognitions should be lower for personally familiar than
for celebrity stimuli. Also, correct recognition of personally familiar stimuli should, on
average, occur at a higher masking level than correct recognition of famous stimuli.

Drawing upon the theory regarding the face recognition advantage (Barsics, 2014),
analogous predictions were made for faces relative to voices. Finally, for the personally
familiar stimuli (i.e., fellow students), we expected the number of shared tutorial groups
of stimulus persons and the participants to be positively associated with recognition
performance of both faces and voices (see Yarmey, Yarmey, Yarmey, & Parliament, 2001).

Method

Participants and Design

Thirty-one second-year psychology students (16 men; age range 19–28 years, \( M = 21.0, \)
\( SD = 1.9 \) ) took part in the experiment in exchange for course credit. The study was
approved by the Ethical Review Committee Psychology and Neuroscience of Maastricht
University (approval number: ECP_2027_27) and follows the rules stated in the Declaration
of Helsinki. Participants were assigned in a 2 (type of familiarity: famous vs. personally
familiar) \( \times 2 \) (modality: face vs. voice) within-participants design.

Stimulus Persons

Ten Dutch celebrities who regularly appear on Dutch television and who are well-known in
the Netherlands (where the study was conducted) were chosen for the set of celebrity stimuli
(see Appendix for a list of the celebrities). Ten second-year students who were well-known in
their year, for example, because they often asked questions during lectures and tutorial
groups or were active in the student association of the faculty served as personally
familiar stimuli. All stimulus persons were Caucasian and the stimuli sets comprised both
males and females. A pilot study yielded that the stimulus persons were also well-known
among the participant population (i.e., second-year psychology students). That is, the
photos of the stimulus persons were presented to the 10 second-year students who also
served as stimulus persons. All stimulus persons (celebrities and students) were correctly
identified by the 10 second-year students.

Audio and Video Fragments

The 10 students’ stimulus fragments were obtained in a room (4 \( \times 2.5 \times 3 \) m), with a carpet
absorbing reverberations. The audio fragments in which the students talked about neutral
subjects (i.e., nothing that could potentially disclose their identities) were recorded with a
Marantz PMD661 Professional solid state digital audio recorder with the internal microphone at a distance of 0.6 m. To avoid a potential reduction of natural variability in the voices during the recording, the students were having a conversation with the assistant in charge of the recording. The visual fragments were recorded with a JVC GR-D820E digital video camera at a distance of 1.5 m. They were naturalistic video clips (see Lander et al., 2001) in which the students were seated, but moved (e.g., head or arm movements), as they were talking. The face and part of the upper body could be seen. The audio fragments were recorded simultaneously with the visual fragments. The audio track was removed from the visual fragments (i.e., the video fragments did not contain any sound).

The 10 celebrities’ fragments were obtained from the Internet. The video and audio fragments came from the same footage. In the video fragments, the background contained no clues regarding the celebrities’ identities and the face and part of the upper body were displayed. In the audio fragments, the celebrities spoke about neutral topics (as established in a pilot study with \( N = 11 \) participants; see Damjanovic & Hanley, 2007, for a similar approach).

All fragments had a duration of 10 to 15 seconds. Four masking levels were applied. The final, or lowest, “masking” level was the unmasked stimulus. Similar to Clark and Foulkes (2007), the audio fragments were masked by lowering the frequency by levels of \(-6, -4,\) and \(-2\) semitones (with \(-6\) being the highest masking level) using Adobe Audition CS6. Visual fragments were masked by pixelating the stimuli at levels of 8, 11, or 14 horizontal pixels per face (i.e., high to low masking) with Adobe Premiere Pro (see Demanet et al., 2007, for a comparable approach). Note that everything that was displayed of the person (i.e., head and upper body area) was pixelated in order to avoid that non-facial cues would reveal the identity of the person (Demanet et al., 2007).

The fragments were integrated into PowerPoint files, with the presentation order counterbalanced across five versions. To avoid priming effects, each of the 20 stimulus persons appeared only once within every version. That is, within a given version, a stimulus person that was presented visually was not presented auditorily and vice versa. Presentation occurred in the following order: five celebrity voices, five student voices, five celebrity faces, and five student faces.

**Procedure**

Participants were informed that the experiment examined the effectiveness of visual and auditory masking. They signed an informed consent form and provided demographic information. Hereafter, one of five PowerPoint presentations was presented, on which participants worked independently. Headphones (Sennheiser HD 25-1) were used to present the audio fragments. Presentation of a given fragment always started with the highest masking level. If the participant reported that he or she did not recognize the person, they were presented with the same fragment at the next lower masking level. This was continued until the participant indicated recognition, or until the final, unmasked fragment had been presented. Following recognition, participants were asked to provide the person’s name or an unambiguous description of the individual if they did not know or could not think of the name.

**Scoring**

Participants’ responses were coded into four categories. “Don’t know” responses were scored when participants indicated that they did not recognize the person even after the
unmasked stimulus had been presented. A correct recognition was scored when the correct name of the person or an unambiguous description of the person (e.g., the king; member of student association that always wears a hoody) were provided. A false recognition was scored when an incorrect name was provided or an unambiguous description indicated a different person (e.g., the Dutch prime minister when in fact the king was presented). A final category, familiar judgments, was employed to consider cases in which participants indicated to recognize the person, but could not provide a name or an unambiguous description.

Results

One third of the famous faces were correctly recognized at the highest masking level (8 pixels; \( M = 34.19\% \), \( SD = 23.77 \)). This significantly deviates from zero, \( t(30) = 8.01, p < .001 \). Of the personally familiar faces, a quarter were correctly recognized at this masking level (\( M = 24.52\% \), \( SD = 21.73 \)), which is also significantly greater than zero, \( t(30) = 6.28, p < .001 \). In contrast, none of the speakers were correctly recognized at the highest masking level (−6 semitones), meaning that recognition was fully eliminated.

To examine the influence of type of familiarity (famous vs. personally familiar) and modality (face vs. voice) on recognition performance, we computed two-way repeated-measure analyses of variance. The average masking level at which correct recognition occurred (1 = highest masking level; 4 = unmasked stimulus), the proportion of correct recognitions (i.e., correct recognitions divided by all recognition decisions) and the proportion of false recognitions (i.e., false recognitions divided by all recognition decisions) served as dependent variables. The proportions were calculated once including the familiarity judgments in and once excluding the familiarity judgments from all recognition decisions. The reason for this approach is that familiarity judgments have been found to be of a different quality than correct or false recognitions (Palmer, Brewer, McKinnon, & Weber, 2010; Yonelinas, 2001). All interactions between type of familiarity and modality were non-significant, \( F_s < 3.24, p_s > .082, \eta^2_s < .10 \), unless otherwise noted.

Recognition Performance as a Function of Type of Familiarity and Modality

Including all stimuli. The means and standard deviations of the proportion of correct and incorrect recognitions and of the average masking level at which correct recognition occurred taking into account all stimuli can be found in Table 1.

Proportion of correct recognitions. As expected, the proportion of correct recognitions (familiarity judgments included) was significantly higher for faces (\( M = 73.87\% \), \( SD = 19.95 \)) than for voices (\( M = 36.13\% \), \( SD = 18.01 \)), \( F(1, 30) = 102.36, p < .001, \eta^2 = 0.77 \). Unexpectedly, the proportion of correct recognitions was significantly higher for famous (\( M = 60.97\% \), \( SD = 18.14 \)) than for personally familiar stimuli (\( M = 49.03\% \), \( SD = 20.06 \)), \( F(1, 30) = 9.83, p = .004, \eta^2 = 0.25 \).

When familiarity judgments were excluded, the proportion of correct recognitions was again significantly higher for faces (\( M = 79.95\% \), \( SD = 20.25 \)) than for voices (\( M = 37.26\% \), \( SD = 18.52 \)), \( F(1, 30) = 119.04, p < .001, \eta^2 = 0.80 \). The main effect of target on the proportion of correct recognitions became nonsignificant, \( F(1, 30) = 2.69, p = .111, \eta^2 = 0.09 \).

Proportion of false recognitions. For the proportion of false recognitions (familiarity judgments included), both main effects were significant, modality: \( F(1, 30) = 5.22, p = .030, \eta^2 = 0.15 \); type of familiarity: \( F(1, 30) = 10.15, p = .003, \eta^2 = 0.25 \), but were qualified by a
significant interaction (see Figure 1), $F(1, 30) = 8.41, p = .007, \eta^2_p = 0.22$. The analysis of the simple main effects yielded that the proportion of false recognitions was higher for famous voices ($M = 19.35\%, \ SD = 18.96$) than for personally familiar voices ($M = 3.87\%, \ SD = 9.55$), $t(30) = 3.86, p = .001, d = 1.04$. No difference was found for famous ($M = 7.74\%, \ SD = 15.21$) versus personally familiar faces ($M = 5.81\%, \ SD = 10.58$), $t(30) = 0.62, p = .540, d = 0.15$. When the familiarity judgments were excluded, analogous results emerged.
The main effect of modality was significant, $F(1, 23) = 80.82, p < .001, \eta_p^2 = 0.78$. On average, faces were correctly recognized at a masking level of $M = 2.20$ ($SD = 0.52$; with 1 = the highest masking level and 4 = the unmasked stimulus) and voices were correctly recognized at a masking level of $M = 3.47$ ($SD = 0.41$).

No differences were found as a function of type of familiarity (famous: $M = 2.84$, $SD = 0.35$; personally familiar: $M = 2.84$, $SD = 0.43$), $F(1, 23) < 0.01, p = .989, \eta_p^2 < 0.01$.

In hindsight, we realized that any significant difference between famous and personally familiar stimuli could be due to the fact that one group was better known to our participants than the other group. Therefore, we conducted a follow-up survey and sent the participants an e-mail with a PowerPoint presentation that contained pictures of the 20 persons who had served as stimuli. Participants were asked to indicate for each of the 20 persons whether they had known them at the time of data collection and if so, to provide their names. Of the $N = 31$ participants, $n = 24$ responded. Results showed that participants knew $M = 9.21$ ($SD = 1.28$) of the 10 celebrities, but only $M = 7.79$ ($SD = 2.21$) of the 10 fellow students, suggesting an advantage for the celebrity stimuli, $t(23) = 2.84, p = .009, d = 0.78$. Therefore, we reran all analyses, taking into consideration only those stimuli that the participants indicated to have known during the time of data collection. The means and standard deviations of the dependent variables taking into account only the stimuli known to participants can be found in Table 2.

### Table 2. Proportion of Correct and False Recognitions and Average Masking Level at Correct Recognition as a Function of Type of Familiarity and Modality (Including Only Stimulus Persons Known to Participants).

|                      | Famous       | Personally familiar | Across target |
|----------------------|--------------|---------------------|---------------|
|                      | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ |
| Proportion of correct recognitions (including familiar judgments) (%) | Voices | 47.25 | 33.31 | 41.59 | 26.57 | 44.42 | 19.76 |
|                      | Faces | 91.38 | 12.84 | 81.59 | 23.03 | 86.49 | 15.35 |
|                      | Across Modality | 69.31 | 17.93 | 61.59 | 19.91 | 65.45 | 15.35 |
| Proportion of correct recognitions (excluding familiar judgments) (%) | Voices | 47.25 | 33.31 | 44.49 | 29.57 | 45.87 | 20.36 |
|                      | Faces | 94.20 | 11.94 | 87.25 | 21.34 | 90.72 | 15.03 |
|                      | Across Modality | 70.72 | 18.91 | 65.87 | 22.29 | 68.30 | 15.67 |
| Proportion of false recognitions (including familiar judgments) (%) | Voices | 16.88 | 16.07 | 2.61 | 9.15 | 9.75 | 9.47 |
|                      | Faces | 3.48 | 9.82 | 6.88 | 12.52 | 5.18 | 9.81 |
|                      | Across Modality | 10.18 | 10.61 | 4.75 | 7.73 | 7.46 | 7.93 |
| Proportion of false recognitions (excluding familiar judgments) (%) | Voices | 16.88 | 16.07 | 2.61 | 9.15 | 9.75 | 9.47 |
|                      | Faces | 3.48 | 9.82 | 7.10 | 12.79 | 5.29 | 9.99 |
|                      | Across Modality | 10.18 | 10.61 | 4.86 | 7.83 | 7.52 | 8.04 |
| Average masking level of correct recognition | Voices | 3.55 | 0.43 | 3.40 | 0.62 | 3.47 | 0.40 |
|                      | Faces | 2.12 | 0.72 | 2.29 | 0.76 | 2.20 | 0.56 |
|                      | Across Modality | 2.83 | 0.34 | 2.84 | 0.47 | 2.84 | 0.31 |

Note. Means sharing the same superscript letter indicates significant main effects with $p < .05$; masking level: 1 = highest masking level, 4 = unmasked stimulus.

Average masking level at correct recognition. The main effect of modality was significant, $F(1, 23) = 80.82, p < .001, \eta_p^2 = 0.78$. On average, faces were correctly recognized at a masking level of $M = 2.20$ ($SD = 0.52$; with 1 = the highest masking level and 4 = the unmasked stimulus) and voices were correctly recognized at a masking level of $M = 3.47$ ($SD = 0.41$).

Analyses including only stimuli known to participants

Proportion of correct recognitions. Again, the proportion of correct recognitions (familiarity judgments included) was significantly higher for faces ($M = 86.49\%$,
SD = 15.35) than for voices (M = 44.42%, SD = 19.76), F(1, 22) = 131.65, p < .001, $\eta^2_p = 0.86$. In contrast to the above analyses and our hypotheses, the main effect of type of familiarity was nonsignificant (famous: M = 61.59%, SD = 19.91), F(1, 22) = 2.78, p = .110, $\eta^2_p = 0.11$. When familiarity judgments were excluded, analogous results emerged.

**Proportion of false recognitions.** For the proportion of false recognitions (familiarity judgments included), while the main effect of modality failed to reach significance, F(1, 22) = 3.98, p = .058, $\eta^2_p = 0.15$, the main effect of type of familiarity was significant, F(1, 22) = 7.30, p = .013, $\eta^2_p = 0.25$. The significant interaction yielded a similar pattern as earlier (see Figure 2), F(1, 22) = 13.82, p = .001, $\eta^2_p = 0.39$: the proportion of false recognitions was higher for the famous (M = 16.88%, SD = 16.07) than personally familiar voices (M = 2.61%, SD = 9.15), t(23) = 4.09, p < .001, d = 1.16. For faces, the proportion of false recognitions did not differ between the famous (M = 3.48%, SD = 9.82) and personally familiar ones (M = 6.88%, SD = 12.52), t(22) = -1.48, p = .152, d = -0.30. The analysis excluding familiar judgments led to analogous results.

**Average masking level at correct recognition.** The main effect of modality was significant, F(1, 17) = 51.34, p < .001, $\eta^2_p = 0.75$. On average, faces were correctly recognized at a masking level of M = 2.20 (SD = 0.56; with 1 = the highest masking level and 4 = the unmasked stimulus) and voices were correctly recognized at a masking level of M = 3.47 (SD = 0.40). No differences were found as a function of type of familiarity (famous: M = 2.83, SD = 0.34; personally familiar: M = 2.84, SD = 0.47), F(1, 17) = 0.01, p = .929, $\eta^2_p < 0.01$.

**Post Hoc Voice Ratings**

To explore possible explanations for the high proportion of false recognitions of the celebrities’ voices, N = 23 (M = 23.4 years, SD = 4.6, age range 18–34 years) Dutch native speakers rated the voices of all 20 stimulus persons and were asked to identify the speakers. The ratings referred to two different aspects of distinctiveness (difficulty to recognize the voice from a group of people and unusualness), as well as attractiveness, memorability, pitch, and presence of an accent. The wording of the items (adapted from Sauerland, Sagana, & Otgaar, 2013), the referring anchors of the employed 7-point Likert scales ranging from 1 to 7, and the means and standard deviations can be found in Table 3.
Table 3. Items for Measuring Difficulty to Recognize, Distinctiveness, Attractiveness, Memorability, Pitch, and Presence of Accent/Dialect, and Post Hoc Ratings of the Celebrities’ and Students’ Voices.

| Item                          | Anchor low (1)                                    | Anchor high (7)                                    | Celebrities  | Students  |
|-------------------------------|--------------------------------------------------|---------------------------------------------------|--------------|-----------|
|                               |                                                  |                                                   | M (SD)       | M (SD)    |
| Difficulty to Recognize*      | How difficult would you find it to recognize this voice from a group of people who are talking to each other? | Very difficult                                    | 4.42 (1.00)  | 3.37 (0.70) |
| Distinctiveness*              | How distinctive or unusual is this voice to you?  | Not distinctive/unusual at all                    | 4.04 (1.14)  | 2.87 (0.67) |
| Attractiveness*               | How attractive is this voice to you?             | Not attractive at all                              | 3.72 (0.69)  | 2.99 (0.86) |
| Memorability*                 | How memorable (i.e., easy to remember) is this voice to you? | Not memorable at all                              | 4.11 (0.93)  | 2.83 (0.68) |
| Pitch                         | How is the pitch of this voice?                  | Very low                                          | 3.62 (0.43)  | 3.72 (0.59) |
| Accent*                       | Does this person speak with an accent/dialect?   | No accent/dialect at all                          | 3.75 (1.16)  | 4.68 (0.83) |

*Significant differences with \( p \leq .001 \).
Compared with the students’ voices, the celebrities’ voices were found to be significantly less difficult to recognize, more unusual, more attractive, and more memorable, \( t(22) \geq 3.72, ps < .001, ds \geq 0.94 \). The students spoke with more accent than the celebrities, \( t(22) = -4.77, p < .001, d = -0.89 \). Pitch ratings did not differ between the celebrity and student speech samples, \( t(22) = -0.82, p = .422, d = -0.19 \). Analogous results emerged when only those stimuli were included in the analyses that the participants of the post hoc study did not recognize. These results suggest that the celebrities’ voices were more distinctive and more special than the students’ voices, independent of actual speaker recognition.

### Table 4. Regression Analyses Predicting the Proportion of Correct and Incorrect Recognitions and the Average Masking Level at Correct Recognition of Voices and Faces from the Number of Shared Tutorial Groups.

| Dependent Variable | \( B \) | \( SE \) | \( \beta \) | \( t \) | \( p \) | \( R^2 \) | \( F \) |
|-------------------|--------|--------|----------|------|------|---------|------|
| **Voices**        |        |        |          |      |      |         |      |
| Proportion of correct recognitions (including familiar judgments) | 0.03  | 0.02  | .30      | 1.71 | .099 | .09     | 2.91 |
| Proportion of correct recognitions (excluding familiar judgments) | 0.03  | 0.02  | .24      | 1.34 | .192 | .06     | 1.78 |
| Proportion of false recognitions (including familiar judgments) | -0.01 | 0.01  | -.33     | -1.89| .068 | .11     | 3.59 |
| Proportion of false recognitions (excluding familiar judgments) | -0.01 | 0.01  | -.33     | -1.89| .068 | .11     | 3.59 |
| Average masking level of correct recognition | -0.03 | 0.05  | -.12     | -0.57| .572 | .01     | 0.33 |
| **Faces**         |        |        |          |      |      |         |      |
| Proportion of correct recognitions (including familiar judgments) | 0.06  | 0.02  | .54      | 3.43 | .002 | .29     | 11.77|
| Proportion of correct recognitions (excluding familiar judgments) | 0.06  | 0.02  | .57      | 3.72 | .001 | .32     | 13.86|
| Proportion of false recognitions (including familiar judgments) | < -0.01| 0.01  | -.02     | -0.11| .917 | <.01    | 0.01 |
| Proportion of false recognitions (excluding familiar judgments) | < -0.01| 0.01  | -.03     | -0.15| .882 | <.01    | 0.02 |
| Average masking level of correct recognition | 0.04  | 0.06  | .14      | 0.77 | .450 | .02     | 0.59 |

Regression of Recognition Performance on the Number of Shared Tutorial Groups

To test the idea that the number of shared tutorial groups predicted recognition performance of the fellow students (i.e., the personally familiar stimuli), we computed regression analyses. Separate regression analyses for the faces and voices were performed because a stimulus person presented in the visual stimulus set was not part of the auditory stimulus set and vice versa, so the number of shared tutorial groups differed for the stimulus persons in the voice and the face stimulus sets. Table 4 displays the results of the regression equations.

The number of shared tutorial groups was not predictive of any of the speaker recognition performance measures. It did predict the proportion of correct recognitions of the faces. No effects were found for the proportion of false recognitions or the level at which correct recognitions of the faces occurred.
Discussion

In the present study, we examined the effects of masking on recognition performance of famous and personally familiar persons' faces and voices, using pixelation and electronic voice disguise. To allow for a fine-grained analysis of the effects of familiarity, the number of shared tutorial groups of the participants and their fellow students were used as a measure of level of familiarity. The lack of masking research on different types of familiarity (Natu & O'Toole, 2011), especially on personally familiar stimuli, is surprising, as the identity of disguised persons particularly needs to be kept secret from personally familiar others. Therefore, our main aim was to compare the effects of masking personally familiar individuals and celebrities.

Before addressing the results concerning the famous and personally familiar stimuli in detail, we should like to point out that we replicated the well-established finding that faces are easier to recognize than voices (Hanley et al., 1998; Yarmey et al., 1994). The proportion of correct recognitions was higher for faces than for voices. Our results indicate that this recognition advantage also occurs when stimuli are masked. From a theoretical perspective, our results seem to suggest that masking does not counteract the mechanisms that have been proposed to account for the face recognition advantage. These include the inhibition of encoding auditory information when simultaneously encoding visual information (McAllister et al., 1993), stronger associations between episodic memory and the visual versus the auditory recognition system (Brédart et al., 1999), and higher expertise in using cues from the visual relative to the auditory domain to extract person information (Barsics, 2014). However, we would also like to highlight that the interpretation of this result is somewhat limited, since the comparison between faces and voices could not be meaningfully made for the dependent variable average masking level at which correct recognition occurred. The different masking levels may not have been directly equivalent for faces and voices. Therefore, this finding requires replication in a study that established the equivalence of the masking levels for faces and voices, before final conclusions can be drawn.

Regarding the comparison between personally familiar and famous stimuli, face processing research has yielded different representations of famous and personally familiar faces (Taylor et al., 2009). That is, face unit activation is increased in personally familiar relative to famous faces (Herzmann et al., 2004). Moreover, personally familiar faces seem to have a wider representation than famous faces, extending beyond the mere formation of visual memory and involving increased activation in brain areas associated with person knowledge (Gobbini et al., 2004). These findings have resulted in the proposition of a face recognition model, in which recognition of a familiar person entails recognition of the visual appearance, as well as the spontaneous activation of person knowledge and an emotional response (see Gobbini & Haxby, 2007, for a review). Regarding speaker recognition, it has been suggested that recognition performance improves when one has engaged in an active conversation with that speaker, because in an active conversation, more information about the voice is attended to and subsequently encoded (Hammersley & Read, 1985). This should put recognition of personally familiar speakers, relative to famous speakers, at an advantage. On the basis of these theoretical considerations, we expected recognition performance to be higher for personally familiar than famous masked stimuli. Our results, however, do not support these hypotheses. Rather, masking was equally effective for famous and personally familiar stimuli: neither the proportion of correct recognitions nor the average masking level at which correct recognition occurred differed as a function of type of familiarity. The initial unexpected finding of a higher proportion of correct recognitions for famous than for
personally familiar stimuli (which was accompanied by an increase in familiarity judgments for the personally familiar stimuli) was an artefact caused by the circumstance that more famous than personally familiar stimuli were known to the participants. The differences disappeared when only those stimuli were considered that were known to participants at the time of testing.

This raises the question of why there were no differences between famous and personally familiar stimuli regarding the proportion of correct recognitions and the average masking level at correct recognition. It is possible that the discussed wider memory representation of personally familiar relative to famous faces (Gobbini et al., 2004; Gobbini & Haxby, 2007) or the difference of activated brain regions (Taylor et al., 2009; Walton & Hills, 2012) do not lead to a recognition advantage of personally familiar stimuli. Moreover, increased face unit activation in personally familiar relative to famous faces (Herzmann et al., 2004), which would suggest an enhanced recognition performance, is not consistently found (Kloth et al., 2006). A second explanation may be that the personally familiar persons employed in the present study were not personally familiar enough to yield significant differences compared with the famous persons. According to the face recognition theory by Gobbini and Haxby (2007), successful recognition of a familiar person involves not only the recognition of the visual appearance but also the activation of person knowledge (e.g., biographical knowledge or personal traits). It may be that the participants only had little person knowledge about their fellow students, which did not differ markedly from the person knowledge about the celebrities. As a consequence, no differences in recognition performance may have occurred. Likewise, it could be that the participants did not build a sufficiently strong active voice memory because they did not actively converse enough with their fellow students (Hammersley & Read, 1985), which would also counteract performance differences. Conducting a follow-up study with best friends or close relatives instead of fellow students as personally familiar stimuli could shed light onto this question because it is to be expected that both person knowledge and active voice memory should be higher for these individuals.

The proportion of false recognitions of the faces was low and did not differ between famous and personally familiar stimuli. Regarding the voices, the proportion of false recognitions was much higher for the famous than for the personally familiar stimuli, even after the stimuli that were unknown to participants had been removed, indicating a recognition advantage of personally familiar over famous speakers. This result could simply be due to the higher familiarity of the personally familiar relative to the famous speakers (Yarmey et al., 2001). Alternatively, participants may just have been more inclined to provide a name, even a wrong one, rather than a “don’t know” response for the famous voices. This may be caused by a possible higher distinctiveness of the celebrities’ voices. The higher distinctiveness could give rise to an, albeit false, feeling of familiarity and confidence to recognize the speaker, which may ultimately lead to a false recognition. To test this explanation, we collected post hoc ratings of the voices regarding distinctiveness and other aspects that may make a voice appear special (e.g., attractiveness). Indeed, the celebrities’ voices were considered to be more distinctive (i.e., more unusual and easier to recognize from a group of people), more memorable, and more attractive than the students’ voices, although absolute values were still around the midpoint of the scales. The accent ratings were an exception to this (the students spoke with more accent) but can be explained by the circumstance that the study was conducted in a part of the Netherlands in which the use of dialect is widespread. Interestingly, the findings indicating higher distinctiveness or specialty of the celebrities’ voices persisted when only the stimuli were analyzed that were not recognized by the participants of the post hoc study. This suggests that the celebrities’ voices
were inherently more distinctive or more special than the students’ voices. Following up on voice distinctiveness, and how this may cause a false feeling of familiarity, would be an interesting venue for future research.

To conduct a more fine-grained analysis of the influence of familiarity on the recognition of masked stimuli, we used the number of shared tutorial groups of the participants and their fellow students who served as stimulus persons as a proxy for personal familiarity. Our hypotheses concerning the association between the number of shared tutorial groups and recognition performance were partly confirmed. The more shared tutorial groups the participants and their fellow students attended, the better their faces were recognized. This is in line with the finding of superior recognition performance when the to-be-identified person is familiar (Johnston & Edwards, 2009). This familiarity advantage did not transfer to the masking level. In other words, more correct recognitions were observed with increasing number of shared tutorial groups, but these did not occur at a higher masking level. This suggests that within the group of personally familiar faces, increased familiarity does not make masking less effective.

In contrast to face recognition, speaker recognition performance was not predicted by the number of shared tutorial groups at all, suggesting that speaker recognition did not improve with increased familiarity. The finding contradicts the results of a study in which identification performance was better for highly familiar speakers (e.g., best friend) than for moderately familiar speakers (e.g., coworker), which in turn were better identified than low familiar (e.g., casual acquaintance) and unfamiliar speakers (Yarmey et al., 2001, Experiment 1). One explanation may be that there was not enough variation of familiarity within the personally familiar stimuli to produce significant effects. However, this consideration is difficult to reconcile with the outcome that the number of shared tutorial groups did significantly predict the proportion of correct recognitions of the faces. Rather, the finding may be an indication of the difficulty to recognize speakers (Barsics, 2014). Two other explanations for the finding are associated with the nature of the tutorial groups. First, the students’ voices are always present together with their faces, which may inhibit encoding of the auditory information (McAllister et al., 1993). Second, because the students do not constantly talk, there is potentially more encoding time for the faces than for the voices. The finding that the level of masking at which correct speaker recognition occurred was not significantly predicted suggests that increased familiarity does not decrease the effectiveness of voice masking measures. To summarize the pattern of results regarding masking and different types of familiarity, we found that the effectiveness of masking differed neither across different types of familiarity (personally familiar vs. famous) nor within the group of personally familiar stimuli. It is important to point out that this study should be considered a first step in investigating the effects of masking and different types of familiarity. Clearly, we encourage replication of our results and strongly urge to conduct future studies on that issue, some of which we have drafted here. The standard deviations of the means we obtained were rather large. This suggests that there are moderators of the effectiveness of masking that need to be identified in future research.

As to the general effectiveness of masking, we had expected that the highest masking level would not fully eliminate recognition. This expectation was confirmed for both famous and personally familiar faces, replicating previous findings (Demanet et al., 2007; Lander et al., 2001). This calls into question the effectiveness of face pixelation to conceal the identity of a person, especially when the stakes are high. In contrast to the faces, none of the speakers was correctly recognized by any of the participants on the highest masking level. This means that a reduction of voice frequency by six semitones is sufficient to fully eliminate speaker recognition. This is in line with Clark and Foulkes (2007) who found that a reduction by
eight semitones reduced performance to chance level, but contradicts Huckvale and Kristiansen (2012) who reported much higher recognition rates on the same masking level. The discrepancy to the present findings is noteworthy because Huckvale and Kristiansen also utilized personally familiar speakers (i.e., fellow students). However, there are important differences in the methodology of the studies. Huckvale and Kristiansen’s masked speakers were chosen from a previously tested pool of speakers based on the criterion that they were recognized the best and fastest. Hence, these speakers were easy to recognize (cf. Clark & Foulkes, 2007). Additionally, the participants of the masking experiment were identical to the ones who had taken part in the study that tested the pool of speakers. Therefore, and unlike our participants, they knew which persons they could potentially expect to identify. Essentially, Huckvale and Kristiansen employed a closed set of stimuli (in contrast to an open set, as used in the present study), making identification easier.

Our study supplements the database on the effectiveness of electronic voice disguise and indicates that a reduction of frequency by six semitones may be an effective measure to eliminate speaker recognition. Given the still small number of studies on this issue, however, it may be premature for a profound recommendation. Until there is a sufficient database, it may be prudent to use other methods to conceal the identity of a speaker. For example, a solution could be that the speaker’s voice is not used at all, but that a different speaker reads out the transcript of the whistleblower’s responses.

Some limitations of the present study should be discussed. First, in hindsight, it is unfortunate that we did not ask the participants at the end of the actual test session, whether they knew all persons presented to them as stimuli. Instead, we enquired about this only after the experiment via e-mail, which we considered the best alternative given the circumstances. Although not all participants responded to our request, we trust the response rate to be sufficiently high to draw valid conclusions given the within-participants design. Second, regarding the personally familiar stimuli and the comparison to famous stimuli, we may have chosen one of the less personally familiar groups, that is, fellow students. In a future study, it should be examined whether the current pattern of results prevails with more personally familiar individuals, such as best friends or close relatives. Third, unlike the source of the students’ fragments, the source of the celebrities’ video and audio fragments was the Internet. Although the celebrities’ fragments were chosen so that they were of acceptable quality, they were not systematically checked for artifacts. Yet, we believe that the results were not influenced by this. Previous research has shown that humans are quite robust speaker and face recognizers, capable of reconstructing degraded or incomplete signals and of recognizing familiar individuals even when video footage is of low quality (Burton et al., 1999; Campbell et al., 2009; Schmidt-Nielsen & Crystal, 2000). If there was some kind of degradation in the fragments, it was of such an extent as would not be problematic for recognition performance. Finally, the number of shared tutorial groups is only a rough proxy for familiarity, because it does not account for how often the students see each other outside of the tutorial groups.

To summarize, the present study yielded first insights into the effectiveness of masking of famous versus personally familiar stimuli. More specifically, masking was found to be equally effective for famous and personally familiar stimuli. Within personally familiar stimuli, level of familiarity did not predict the effectiveness of masking. This suggests that any differences in memory representation of personally familiar and famous stimuli may not lead to enhanced recognition performance. Concerning the overall effects of masking, while the highest masking level applied in our study eliminated voice recognition, this was not the case for face recognition. As a consequence, pixelating a face to conceal a person’s identity cannot be recommended.
Acknowledgements
We would like to thank Myrr van den Broek, Annelot van Coberen, Fleur van Dijk, Jonas Kretzschmar, Lisa Schoenmakers, and Andrea Wolfs for data collection, and Berber Jacquemijns, Shanice Janssens, Joyce Mertens, Marly Schmitz, Paola Smit, Roan-Paul Spölmink, Jowé Waucomont, Raf Widdershoven, Romy Zanders, and Fenny Zwambach for contributing to the stimulus set.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes
1. There are studies comparing face and voice recognition, in which only the faces but not the voices were masked in order to reduce face recognition performance to voice recognition performance (Brédart, Barsics, & Hanley, 2009; Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009; Hanley & Turner, 2000). Rather than studying the effects of masking, however, the research aim of these studies was to examine the mechanism behind the finding that faces are more easily recognized than voices.
2. Originally, we had aimed for −8 semitones as the highest masking level (as in Clark & Foulkes, 2007), but the male voices became incomprehensible at this level.
3. All data are available from the first author.
4. Note that the proportion of false recognitions is not the inverse of the proportion of correct recognitions, because all recognition decisions in the denominator additionally include the “don’t know” responses and, depending on the analyses, the familiarity judgments.
5. We refrained from interpreting this mean difference in terms of a better recognition performance. It may be that the disguise at a specific level (e.g., 8 horizontal pixels vs. −6 semitones on the highest masking level) was not directly equivalent for faces and voices, rendering a comparison across modalities difficult for this dependent variable.

References
Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? European Journal of Cognitive Psychology, 3, 87–103. doi:10.1080/09541449108406221
Barsics, C. (2014). Person recognition is easier from faces than from voices. Psychologica Belgica, 54, 244–254. doi:10.5334/pb.ap
Bhatia, S. K., Lakshminarayanan, V., Samal, A., & Welland, G. V. (1995). Human face perception in degraded images. Journal of Visual Communication and Image Representation, 6, 280–295. doi:10.1006/jvci.1995.1024
Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar face matching. Applied Cognitive Psychology, 27, 707–717. doi:10.1002/acp.2970
Brédart, S., Barsics, C., & Hanley, R. (2009). Recalling semantic information about personally known faces and voices. European Journal of Cognitive Psychology, 21, 1013–1021. doi:10.1080/09541440802591821
Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*, 243–248. doi:10.1111/1467-9280.00144

Campbell, J. P., Shen, W., Campbell, W. M., Schwartz, R., Bonastre, J.-F., & Matrouf, D. (2009). Forensic speaker identification. A need for caution. *IEEE Signal Processing Magazine, 26*, 95–103. doi:10.1109/MSP.2008.931100

Clark, J., & Foulkes, P. (2007). Identification of voices in disguised speech. *International Journal of Speech, Language and the Law, 14*, 195–221. doi:10.1558/itsll.v14i2.195

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantization effects in face recognition. *Perception, 23*, 129–146. doi:10.1068/p230129

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics, 58*, 602–612. doi:10.3758/bf03213093

Damjanovic, L., & Hanley, J. R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory & Cognition, 35*, 1205–1210. doi:10.3758/bf03193594

Demanet, J., Dhont, K., Notebaert, L., Pattyn, S., & Vandierendonck, A. (2007). Pixelating familiar people in the media: Should masking be taken at face value? *Psychologica Belgica, 47*, 261–276. doi:10.5334/pb-47-4-261

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*, 32–41. doi:10.1016/j.neuropsychologia.2006.04.015

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage, 22*, 1628–1635. doi:10.1016/j.neuroimage.2004.03.049

Hammersley, R., & Read, J. D. (1985). The effect of participation in a conversation on recognition and identification of the speakers’ voices. *Law and Human Behavior, 9*, 71–81. doi:10.1007/BF01044290

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, 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. *The Quarterly Journal of Experimental Psychology Section A, 51*, 179–195. doi:10.1080/713755751

Hanley, J. R., & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *The Quarterly Journal of Experimental Psychology Section A, 53*, 1105–1116. doi:10.1080/02724980050156317

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science, 180*, 1194–1197. doi:10.1126/science.180.4091.1194

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What’s special about personally familiar faces? A multimodal approach. *Psychophysiology, 41*, 688–701. doi:10.1111/j.1469-8864.2004.00196.x

Huckvale, M., & Kristiansen, A. L. (2012, June). Effectiveness of electronic voice disguise between friends. Paper presented at the 46th Audio Engineering Society Conference on Audio Forensics, Denver, CO.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*, 577–596. doi:10.1080/09658210902976969

Kloth, N., Dobel, C., Schweinberger, S. R., Zwitserlood, P., Bolte, J., & Junghofer, M. (2006). Effects of personal familiarity on early neuromagnetic correlates of face perception. *European Journal of Neuroscience, 24*, 3317–3321. doi:10.1111/j.1460-9568.2006.05211.x

Künzel, H. J. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics, 7*, 149–179. doi:10.1558/sll.2000.7.2.149

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology, 15*, 101–116. doi:10.1002/1099-0720(200101/02)15:1<101::AID-ACP697>3.0.CO;2-7

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 identifications. *Basic and Applied Social Psychology, 14*, 161–170. doi:10.1207/s15324834basp1402_3
Natu, V., & O’Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology, 102*, 726–747. doi:10.1111/j.2044-8295.2011.02053.x

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica, 133*, 137–145. doi:10.1016/j.actpsy.2009.11.002

Reich, A. R., & Duke, J. E. (1979). Effects of selected vocal disguises upon speaker identification by listening. *The Journal of the Acoustical Society of America, 66*, 1023–1028. doi:10.1121/1.383321

Sauerland, M., Sagana, A., & Otgaar, H. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology, 18*, 371–381. doi:10.1111/j.2044-8333.2012.02049.x

Schmidt-Nielsen, A., & Crystal, T. H. (2000). Speaker verification by human listeners: Experiments comparing human and machine performance using the NIST 1998 speaker evaluation data. *Digital Signal Processing, 10*, 249–266. doi:10.1006/dsp.1999.0356

Taylor, M. J., Arsalidou, M., Bayless, S. J., Morris, D., Evans, J. W., & Barbeau, E. J. (2009). Neural correlates of personally familiar faces: Parents, partner and own faces. *Human Brain Mapping, 30*, 2008–2020. doi:10.1002/hbm.20646

Walton, B. R. P., & Hills, P. J. (2012). Face distortion aftereffects in personally familiar, famous, and unfamiliar faces. *Frontiers in Psychology, 3*, 1–8. doi:10.3389/fpsyg.2012.00258

Yarmey, A. D. (2007). The psychology of speaker identification and earwitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology* (pp. 101–136). Mahwah, NJ: Lawrence Erlbaum Associates.

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8*, 453–464. doi:10.1002/acp.2350080504

Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology, 15*, 283–299. doi:10.1002/acp.702

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*, 361–379. doi:10.1037/0096-3445.130.3.361.

---

**Appendix: List of Celebrities Included in the Experiment**

Carlo Boszhard (TV presenter)

Herman den Blijker (cook and TV presenter)

Nicolette van Dam (actress)

Carice van Houten (actress)

Chantal Janzen (actress)

Gerard Joling (singer)

Paul de Leeuw (TV presenter)

Willem-Alexander van Oranje-Nassau (King of the Netherlands)

Yolanthe Sneijder-Cabau (actress)

Albert Verlinde (TV presenter)