Article

Social Psychological and
Personality Science
1–11

DOI: 10.1177/19485506251407072
journals.sagepub.com/home/spp
Sage

Facial Impressions Are Not Set in Stone: Voice-Driven Changes in Facial Representations Obtained via Reverse Correlation

Matteo Masi[1] and Marco Brambilla[1]

Abstract

Research has shown that face-based impressions change when integrated with valence-inconsistent voices. Yet, it remains unclear whether listening to a voice also alters mental representations of faces. We tested this possibility using reverse correlation (RC) in two preregistered experiments with Italian Prolific users. In Experiment 1 (N = 61), participants formed impressions based on negative behavioral statements about a target person (Time 1) and completed the RC task to visualize face representations (classification images). Next, they formed new impressions on valence-(in)consistent voices (Time 2) before repeating the RC task. In Experiment 2 (N = 120), participants memorized a positive or negative face at Time 1 to measure spontaneous face impressions without requiring an explicit judgment. Classification images were created only after listening to valence-(in)consistent voices at Time 2. Across both experiments, independent judges rated the images (N_{total} = 140). Valence-inconsistent voices changed the perceived valence of the mental face representations. Implications for cross-modal impression formation are discussed.

Keywords

impression updating, face, voice, reverse correlation, cross-modal integration

Handling Editor: Adam Hahn

First impressions are not immutable. It would be maladaptive if individuals always clung to initial perceptions despite subsequent, conflicting information (McArthur & Baron, 1983). Recent research indicates that people integrate cross-modal cues—such as behavioral, facial, and vocal information—when forming impressions of unknown others (Masi et al., 2024; Shen et al., 2020, 2023). Indeed, impressions based on facial cues are updated when valence-inconsistent vocal information is subsequentially heard (Masi et al., 2022). While previous studies have almost exclusively relied on self-report measures to capture such changes, no prior work has tested whether hearing a voice can influence how a person's face is mentally represented. Here, we tested this possibility by using the reverse correlation approach (Dotsch et al., 2013).

Cross-Modal Impressions and Sequential Cue Integration

Social perception is often multimodal. People routinely form judgments based on various cues, including observed behaviors (e.g., "Mark donated money to charity"), facial features, and vocal tones (Freeman et al., 2020; McAleer & Belin, 2018; Sutherland & Young, 2022). Traditionally, these cues have been studied in isolation, and most attention focused on behaviors and faces (Skowronski et al., 2008; Todorov et al., 2025). However, voices are also key cues for impression formation (Farley, 2025).

When faces and voices are presented simultaneously, face impressions are often shaped by vocal information and vice versa, resulting in integrated impressions (Masi et al., 2023; Mileva, 2025; Rezlescu et al., 2015; Vives et al., 2023). Sometimes, however, facial and vocal cues are perceived sequentially, with the second cue leading to impression updating. For example, face-based impressions can be updated by valence-inconsistent voices, resulting in final impressions shifted in valence (Masi et al., 2022). Moreover, reading negative behavioral information can create an impression that may later be revised by a positive voice or face (Masi et al., 2024). Notably, impressions

[1]University of Milano–Bicocca, Italy

Corresponding Author:
Matteo Masi, University of Milano–Bicocca, Building U6, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy.
Email: matteo.masi@unimib.it

formed from negative cues, particularly faces, appear more susceptible to updating when followed by a positive voice. As such, the new information may be perceived more relevant or boost confidence about one's impressions more than the initial cue (Cone et al., 2017).

Although previous research advanced our understanding of how social cues are integrated over time, it has primarily relied on self-report or implicit measures (e.g., AMP; Shen et al., 2020) assessed immediately after exposure. To our knowledge, no prior study has examined whether listening to a voice can alter the face representation of an unknown other. This raises the question of how auditory cues might influence the integrated impression of a person—reflected not only in judgments but also in the way their face is mentally visualized. Drawing on evidence for overlapping processing of facial and vocal cues (Kuhn et al., 2017; Lavan, 2023; Young et al., 2020) and on theoretical models of multimodal person perception (Bülthoff & Newell, 2017; Freeman et al., 2020; Mileva, 2025), we propose that facial and vocal information are integrated into a unified mental representation of a person. To examine how this integration might affect visual representations, we used sequential presentation—where the voice followed the face—and applied reverse correlation (RC; Dotsch et al., 2013), a technique that generates visual proxies of internal representations.

How to Visualize Representations

RC was introduced as a method to generate visual proxies of mental representations and has been used to examine how prototypical members of social categories (e.g., professions, ethnic groups, or religious groups) and traits (e.g., trustworthiness and dominance) are represented (Brinkman et al., 2017). Participants are usually shown faces overlaid with random noise and asked to select the ones that resemble the typical category member (e.g., an untrustworthy person). These choices, along with their noise patterns, are used to construct classification images, which reveals features guiding judgment and approximates the underlying mental representation. Such images are then rated by independent judges on social dimensions to quantify visual bias (Brinkman et al., 2017).

RC has been applied in two main ways. In one, participants never see the original face but receive indirect information through behavioral descriptions or labels. For example, Klein et al. (2021) asked participants to learn about a target's behaviors and then generate images of their mental representations. Even without ever seeing the target's face, participants selected noisy faces reflecting their internal representation before and after counter-attitudinal information. This approach captures shifts in representations but based solely on behavioral cues, ignoring cross-modal integration. In the second variation, participants see target faces and then generate images following manipulation. Rougier and De Houwer (2025) found that neutral faces paired with positive or negative stimuli produced more positive or negative representations, respectively. While consistent with our reasoning, their study used neutral faces and simultaneous pairings, not sequential valence-inconsistent voices. Thus, no previous study has directly examined how voices heard after a face can reshape its mental representation.

The Present Research

The present work aimed to extend research on cross-modal impression formation by testing whether valence-inconsistent auditory information (i.e., a voice) can shape how a person's face is mentally represented, as inferred through reverse correlation (RC; Klein et al., 2021; Rougier & De Houwer, 2025). While previous studies have shown that voices influence face-based impressions (Masi et al., 2022), we asked whether such effects also alter participants' internal representations of facial appearance. RC complements explicit measures by offering a bottom-up, data-driven proxy of cross-modally informed representations, without assuming how such representations appear in the mind (Dotsch & Todorov, 2012). Rather than focusing on judgments about who a person is, we investigated how the integration of cross-modal cues reshapes the mental image of their face (Klein et al., 2021).

In Experiment 1, we tested updating of face representations when they were formed from reading behaviors (Klein et al., 2021). Participants first received negative behavioral information at Time 1 (T1) and then heard the target's voice—either valence-consistent or valence-inconsistent—at Time 2 (T2). Both impressions and representations were measured at T1 and T2, to assess changes across voice conditions.

An advantage of RC is that it allows us to detect changes in face representations without instructing participants to form and explicit impressions. Thus, in Experiment 2, we tested whether representations spontaneously shift due to hearing the voice when the facial information came from a picture (Rougier & De Houwer, 2025). Participants were initially presented with the target's positive or negative face for a memory task, immediately followed by the voice (either valence-consistent or inconsistent), before measuring face representations. In this experiment, only T2 representations were compared, as the original face was displayed at T1 making it pointless to measure representations right after. Notably, we did not ask to form an impression of the target either for the face or for the voice, thereby providing first evidence of an unconstrained, voice-driven change of face representations.[1]

Experiment 1

Experiment 1 tested whether the face representation of a person formed from negative behavioral information (T1) would update after participants listened to vocal information that was either inconsistent or consistent in valence with the initial information (T2).

Method

The study employed a 2 (voice type: consistent vs. inconsistent) × 2 (time: T1 vs. T2) mixed design, with the first factor varying between-subjects, at the images production phase, and fully within-subjects at the images rating phase. As in standard RC, visual representations were generated by (1) superimposing random noise patterns onto a shared base face, (2) averaging the noise from selected faces across trials, and (3) overlaying the averaged noise back onto the base face to create the classification images.

Transparency and Openness. We obtained approval from the University Ethics Committee. For both experiments, we reported how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. All data, analysis scripts in *R* (version 4.4.2), and materials are available at https://osf.io/ch346. Reverse correlation was conducted using the R packages *rcicr* (Dotsch, 2023) and *schmitz* (Schmitz, 2025).

The experiments were preregistered, and all deviations documented (Experiment 1: https://aspredicted.org/dvbs-sfw9.pdf; Experiment 2: https://aspredicted.org/2rvw-zz3k.pdf).

Power Analyses and Sample Size. For the images production phase, there was no proper way to estimate the sample size. Previous studies suggested collecting about 30 producers per condition, thus 60 in total (Schmitz et al., 2024 Study 1). As images were created in Part 1 and rated in Part 2, power needs to be discussed in relation to both parts. A simulation (shiny.ieis.tue.nl/anova_power/) using data from a similar study on impression formation without RC (Masi et al., 2022, Experiment 2) showed that 60 participants ensure 83% of power to find an interaction effect of *b* (unstandardized difference) = .14 (scale: 1–7) in a 2 × 2 mixed design, which suggests that our study is well-powered for the range of interaction effects we observed across Parts 1 and 2 (*b*s > .28).

For the rating phase, we could not estimate the fixed and the random effects due to the novelty of the research question. Thus, in line with previous studies (Dotsch et al., 2008, 2013; Schmitz et al., 2024; Study 1), we collected 70 participants. We then conducted a safeguard sensitivity analysis (Perugini et al., 2014) with the R package *simr* (Green & MacLeod, 2016). We simulated the power using the 20% confidence interval (CI) of the interaction effect (*b* = .26) of the images ratings, which showed that 70 participants had more than 99% of power to observe it (alpha = .05).

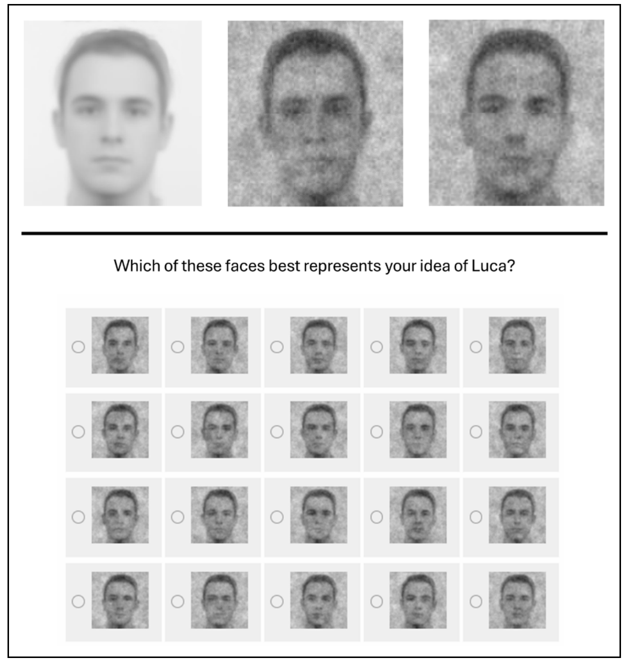

Figure 1. *The Base Face, Taken From Schmitz et al. (2024) (Left), With One Pair of Its Oriented (Center) and Inverted (Right) Noisy Versions, and an Example of a Single Trial From the Brief-RC Procedure (Below).*

Part 1: Classification Images Production

Participants. We collected 61 Italian Prolific users (24 female, M_{age} = 37.72; SD_{age} = 11.80). We checked too fast responses in the RC task (<200 ms for >95% of trials) for excluding extremely inattentive participants, but there were none.[2]

RC Approach. We followed the brief RC method, which compared with the standard RC, reduces trials significantly (36 vs. 360) by increasing the number of face pictures per trial (20 vs. 2) without compromising performance (Schmitz et al., 2024).

To generate faces for the RC task, noise patterns were superimposed onto a shared base face to create different images. We created 720 images (150 × 150 pixels) from a single picture of a Caucasian male individual obtained from the average of six face models with neutral expressions extracted from the Radbound Face Database (Langner et al., 2010; the base face was taken from Schmitz et al., 2024, Study 1). The 720 faces were in grayscale, and half were obtained by adding (oriented image) and half by subtracting (inverted image) random sinusoidal noise patterns from the base face.

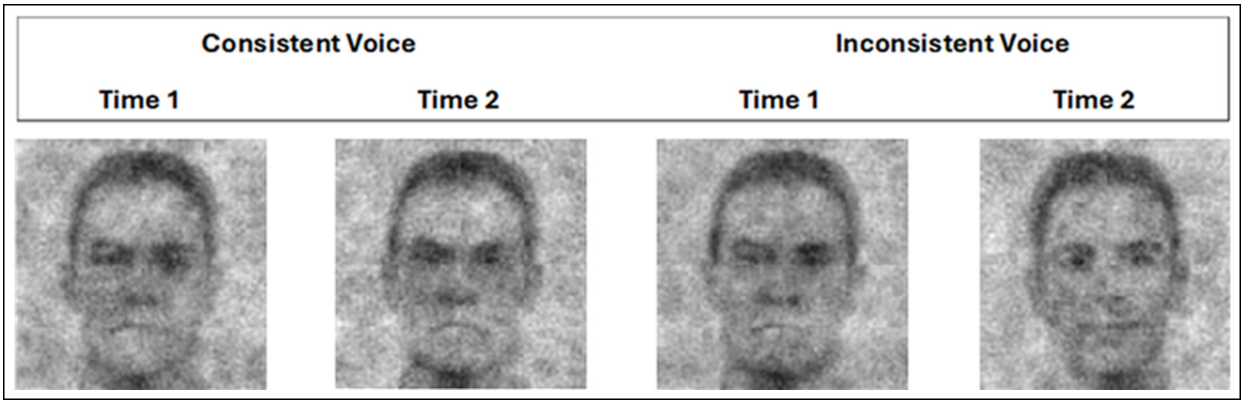

Figure 2. *Examples of Subgroup Classification Images Generated as a Function of Time and Voice Consistency.*
Note. Classification Images at Time 1 were generated from negative behaviors.

In the RC task, participants were presented with 20 of these pictures (without replacement) presented in a 4 × 5 matrix for 36 trials. The task consisted of selecting the single face that better represented their own idea of the target person. See Figure 1 for examples.

Noise patterns from participants' selections were averaged and overlaid on the base face to create classification images, approximating their mental representations. To balance quality and variability, we avoided individual-level images, which lacked definition due to the small number of trials of the brief RC. We also did not limit ourselves to condition-level images, which could obscure individual differences and inflate Type I error (Cone et al., 2021a). Instead, we generated 30 subgroup images per condition, 120 in total, by randomly sampling five participants per iteration without replacement within iterations, but with replacement across iterations (Rougier & De Houwer, 2025). This method preserved individual variability while improving signal quality (see Figure 2).

Materials and Procedure. We used Qualtrics (Provo, UT) and combined the impression updating procedure with RC. Participants were introduced to an impression formation task. We asked them to provide their demographic information and then to read the instructions carefully. They were presented with a target, Luca, and with six behaviors that he has performed in the past, one at a time and in random order. These behaviors were randomly selected from a pool of 20 behaviors that were pretested (24 participants, 20 females, M_{age} = 23.16, SD_{age} = 2.39) to be morally negative on a scale from 1 to 7 (M = 2.85, SD = .275, see Masi et al., 2024, and Supplemental Material [SM]). They were then asked to provide their impression of Luca (1: *negative*, 7: *positive*) and to participate in a "decision task," that was the RC task.

Participants were introduced to the second part. They were asked to listen to six recordings of one of two Italian male speakers in random order (taken from Masi et al., 2022). The sentences they uttered were neutral in content, and each lasted 3 s (e.g., "I sat on the chair," see SM). The original tone of the speakers was as neutral as possible, to avoid the interference of emotional intonation, but voices were artificially modified in pitch to create positive and negative voice-based impressions (McAleer et al., 2014); eventually, the producers listened to six examples of low-pitched or high-pitched voices from one of the two speakers. A pretest (57 participants, M_{age} = 25.19, SD_{age} = 8.14, 35 females) demonstrated that low-pitched voices were perceived negatively (scale −3: *negative* to + 3: *positive*, $M_A = -0.40$, $SD_A = 1.61$; $M_B = -0.42$, $SD_B = 1.36$), and high-pitched voices positively ($M_{\text{A}} = 0.75$, $SD_A = 1.33$; $M_B = 0.56$, $SD_B = 1.54$), with a nonsignificant difference in valence extremity (see SM). Participants were asked to focus on the vocal characteristics. They then provided their impression of Luca once again (1: *negative*, 7: *positive*) before doing the RC task a second time, with the noisy pictures presented in a different random order. They were debriefed and thanked.

Part 2: Classification Images Ratings

Participants. We collected 70 Italian Prolific users (34 female and 36 male, M_{age} = 35.26; SD_{age} = 10.53). We checked who provided identical responses to more than 95% of the images, but none did.

Procedure. Participants were presented with an impression formation task. They were told that they would see faces of different people in grayscale and slightly blurred to make the task more difficult. They were first presented with the four condition-level images, and then with a sample of 40 subgroup images (10 per condition), one at a time and in random order. We asked for their impression of the faces (1: *negative*, 7: *positive*). They were debriefed and thanked.

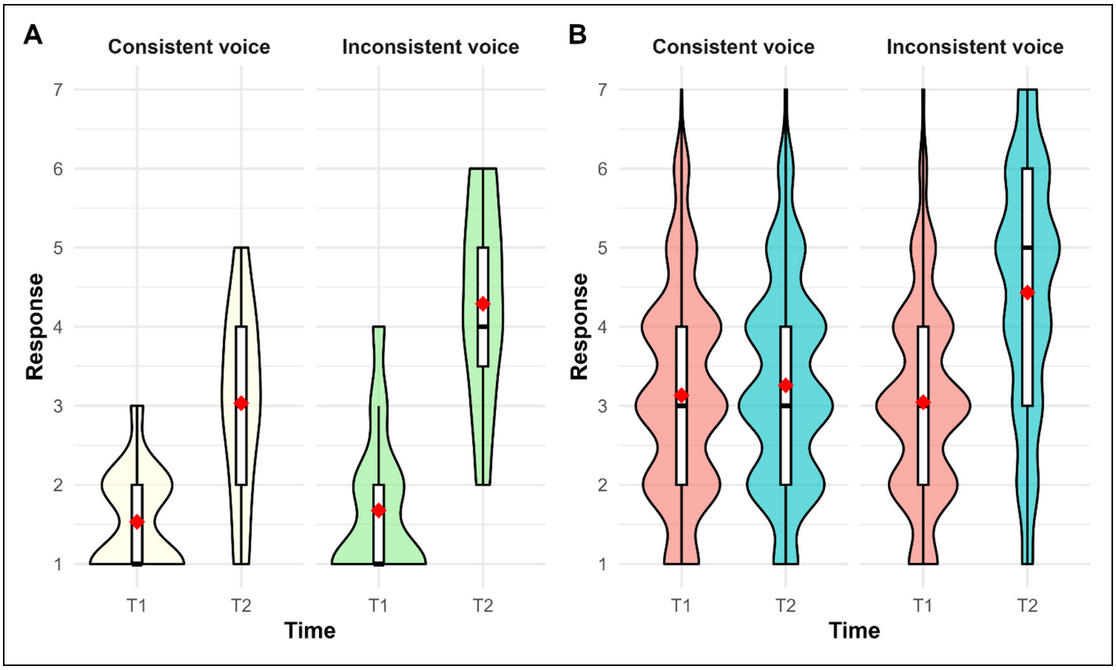

Figure 3. *Violin Plots Showing Ratings of (A) the Target Person Provided by Classification Images Producers and (B) the Subgroup Classification Images Rated by Independent Raters, as a Function of Time and Voice Consistency.*
Note. The red dot indicates the mean.

Results and Discussion

In all analyses, we used mixed-effects models when possible (R package *lmerTest*, Kuznetsova et al., 2017) and simple effects for decomposing interactions (R package *emmeans*, Lenth, 2022). We simplified the random effects structure by starting from the maximal theoretical model and removing terms stepwise until convergence and singularity warnings were resolved (Matuschek et al., 2017). The steps and the rationale are reported in SM. We reported unstandardized effects and 95% CIs as there is no agreement on the proper way to calculate standardized effect sizes for mixed-effects models.

Impressions. We first analyzed whether the producers updated their self-reported negative impressions in a mixed-effects model with random intercepts for participant ID and fixed effects for voice type and time, and their interaction. We found a significant main effect of time, $b = -1.02$, 95% CI = [−1.18, −.88], $t(59) = -13.39$, $p < .001$, meaning that impressions at T2 were more positive than those at T1, and a significant effect of voice type, $b = -.35$, 95% CI = [−.56, −.14], $t(59) = -3.30$, $p < .001$, meaning that the average impression created by an inconsistent voice was more positive than that due to a consistent voice. These effects were qualified by a significant interaction effect, $b = .28$, 95% CI = [.13, .43], $t(59) = 3.62$, $p < .001$. The interaction showed that, despite increasing in positivity, the difference between T1 and T2's impressions was higher when the voice was positive, $b = -2.61$, 95% CI = [−3.04, −2.18], $t(59) = -12.31$, $p < .001$, rather than negative, $b = -1.50$, 95% CI = [−1.94, −1.06], $t(59) = -6.85$, $p < .001$, which we interpret as a greater impression updating when the voice was inconsistent (see Figure 3).

Classification Images Ratings. We proceeded in analyzing the subgroup images ratings in a mixed-effects model accounting for the fixed effects of condition and time, and their interaction, and random slopes for voice condition by participant ID and random intercept for stimuli ID (the analysis on condition-level images gave identical results, see SM). We found significant effects of time, $b = -.37$, 95% CI = [−.49, −.27], $t(115) = -6.81$, $p < .001$, showing that impressions of the images changed from T1 to T2, and of voice type, $b = -.27$, 95% CI = [−.39, −.16], $t(115) = -4.66$, $p < .001$, meaning that, on average, the images created when listening to the positive voice were judged more positively than those based on a negative voice. The two main effects were qualified by a significant interaction effect, $b = .32$, 95% CI = [.21, .42], $t(115) = 5.67$, $p < .001$. Crucially, it showed that the difference in rating between images of the face representations at T1 and T2 following the positive voice, $b = -1.39$, 95% CI = [−1.70, −1.08], $t(115) = -8.83$, $p < .001$, was higher than that following the negative voice, which was not significant, $b =$

$-.13$, 95% CI = $[-.44, .18]$, $t(115) = -.81$, $p = .419$ (see Figure 3).

Thus, we found that inconsistent vocal information can update the face representation formed by negative behavioral information. Specifically, participants who initially evaluated a target person negatively updated their face representation toward a more positive direction after hearing a positive voice rather than a negative voice. Results were consistent with the impressions reported by the images producers, even though the latter showed signs of dilution, as negative impressions became more positive after adding the negative voice, which may be less diagnostic than behaviors (Nisbett et al., 1981). Thus, both impressions and representations were integrated with and updated by cross-modal, vocal cues.

In this experiment, we introduced behavioral information of the target at the first stage, after which we measured the inferred face representation. However, this representation was influenced by individual mental models of a negative person. While we confirmed that participants' face representations inferred from behavioral cues changed in response to new, positive vocal information, we had limited control over *how* the target looked. Moreover, we investigated only the negative-to-positive direction of change, providing no answer for the reversed direction. For this reason, In Experiment 2, we presented a face directly at T1 and considered the positive-to-negative direction.

Experiment 2

In Experiment 2, we aimed to control the initial visual exposure to the face. Rather than presenting behavioral information at T1, participants were shown a positive or negative face. Then, they were exposed to the target's voice, either valence-consistent or inconsistent. RC was subsequently applied to noisy images generated from the shown face. This approach allowed us to assess whether inconsistent voices alter face representations based on an actual face more than consistent voices. Previous research found negative face-to-positive voice impressions changing more than the opposite valence order (Masi et al., 2022), so we investigated this possibility.

Also, Experiment 1 required participants to provide explicit impressions at both T1 and T2 before performing the RC task—a procedure that might have enhanced the likelihood of detecting an RC effect. In Experiment 2, by omitting the explicit judgments before the RC task and asking for deciding on the face more similar to that of the target instead of the one corresponding to their idea of the target, we sought to capture a less constrained influence of the voice. Thus, if classification images varied by vocal condition, it would suggest that participants' face representations were shaped by the integrated voice, even without explicit instructions to form an impression.

Method

The design was a 2 (voice: consistent vs. inconsistent) $\times$ 2 (order: negative face first vs. positive face first) between-subjects design,[3] which becomes fully within-subjects for the images ratings part.

Power Analyses and Sample Size. As in the first experiment, we collected 30 participants per condition, thus 120 in total. A simulation (shiny.ieis.tue.nl/anova_power/) for an interaction effect of $b = .12$ (from Masi et al., 2022, Experiment 2), accounting for the different final impressions after seeing positive or negative faces and hearing (in)consistent voices, gave 97% of power to observe this effect (alpha = .05).

For the images rating phase, we ran an a priori simulations using *simr* for a conservative difference between T2 judgments (consistent vs. inconsistent voice) based on the lower 95% CI ($b = .44$) of Experiment 1. Since in this experiment we also test valence order differences, we ran a simulation including a voice consistency and face valence interaction of $b = .12$ (from Masi et al., 2022)[4] and random intercept per participant and stimulus ID. In both, we found that 70 participants provided 99% power (alpha = .05). Since the final model included a random slope for face valence per participant ID, which we did not consider a priori, a safeguard analysis including these parameters and using the 20% CI of the observed interaction effect ($b = .098$) showed that we had more than 99% of power to observe smaller effects.

Part 1: Classification Images Production

Participants. We collected data from 120 Italian Prolific users (60 females, $M_{age} = 35.40$; $SD_{age} = 9.93$).

Materials and Procedure. To mask the goal, participants were introduced to a memory task. After collecting the demographics, they were shown instructions. They were asked to meet a person, Diego, and look with attention at his face for about 9 s. The picture was an artificial average created by joining nine perceptually negative (we used trustworthiness as it strongly correlates with valence, Todorov et al., 2015; normative data, $M = 2.68$, $SD = .19$) or positive ($M = 3.39$, $SD = .30$) faces taken from the Chicago Face Database (Ma et al., 2015), in grayscale and slightly blurred (see Figure 4). Then, they were asked to listen for about 9 s to three positive or negative voice samples randomly taken from the pool of Experiment 1, paying attention to the voice characteristics, before doing the RC task.

The RC task was based on the 720 noisy pictures generated from the face presented at T1. In each of the 36 trials, they were asked to select the face among the 20 pictures that was most similar to Diego. Eventually, they provided

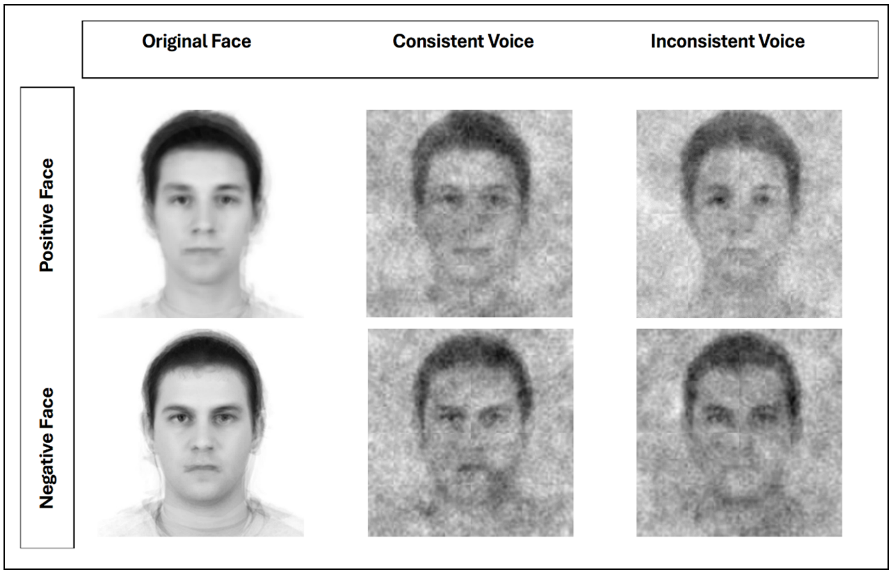

Figure 4. *Original Positive and Negative Faces (left) and Example Subgroup Classification Images Generated After Participants Heard Consistent or Inconsistent Voices Following Face Presentation (Center and Right).*

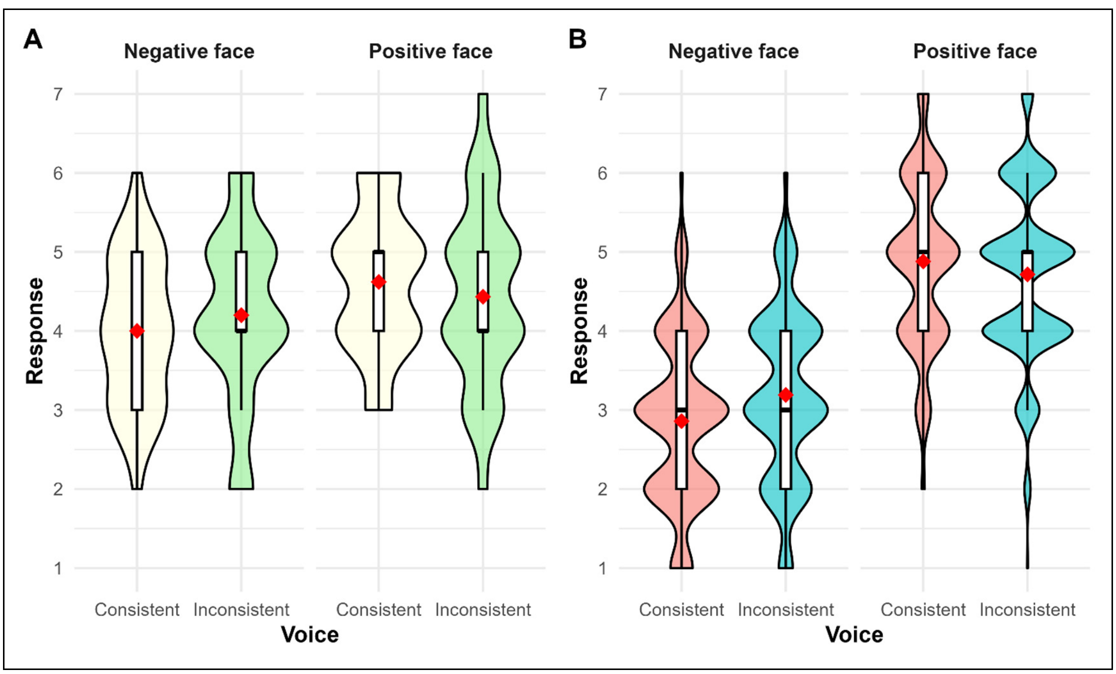

Figure 5. *Violin Plots Showing Ratings of (A) the Target Person Provided by Classification Images Producers and (B) the Subgroup Classification Images Rated by Independent Raters, as a Function of Face Valence and Voice Consistency.*
Note. The red dot indicates the mean.

an explicit judgment about Diego (1: *negative*, 7: *positive*) and to what extent they were surprised by the voice of the person before being debriefed.[5]

We then created the condition-level and subgroup images as in Experiment 1 (see Figure 4).

Part 2: Classification Images Ratings

Participants. We collected 72 Italian Prolific users, but were left with 70 participants (32 female and 36 male, 2 nonbinary; M_{age} = 38.2, SD_{age} = 13.06) when applying exclusion criteria (more than 95% of the responses were identical).

Procedure. The procedure followed that of Experiment 1.

Results and Discussion

Impressions. The data were analyzed using a mixed-effects model that included face valence, voice consistency, and their interaction. Results indicated that impressions of Diego were more positive when the face was originally positive rather than negative, $t(116) = -2.29$, $p = .029$, Cohen's $d = -.43$, 95% CI = [−.79, −.06]. However, there was no significant difference between impressions of Diego with consistent versus inconsistent voices, $t(116) = -.03$, $p = .973$, Cohen's $d = .005$, 95% CI = [−.36, .36], and the interaction was not significant, $t(116) = -.10$, $p = .30$, Cohen's $d = .02$, 95% CI = [−.35, .38] (see Figure 5).

Classification Images Ratings. We analyzed the ratings using a mixed-effects model that included fixed effects for face valence, voice consistency, and interaction, along with random slopes for face valence by participant ID and a random intercept for stimulus ID (results based on the condition-level images were identical, see SM).

We found a significant effect of face valence, $b = -0.89$, 95% CI = [−1.00, .77], $t(114.46) = -29.61$, $p < .001$, indicating that images generated from positive faces were evaluated more positively than those generated from negative faces. There was no significant main effect of voice consistency, $b = -0.04$, 95% CI = [−.10, .02], $t(114.46) = -1.40$, $p = .160$. These effects were qualified by a significant interaction, $b = -0.12$, 95% CI = [−.18, −.06], $t(114.46) = -4.15$, $p < .001$. Simple effects revealed that images based on a negative face paired with an inconsistent voice were evaluated more positively than images based on pairing with a consistent voice, $b = -0.33$, 95% CI = [−.50, −.16], $t(114) = -3.93$, $p < .001$. In addition, images generated from a positive face paired with an inconsistent voice were evaluated as descriptively more negative than images generated from pairing with a consistent voice, $b = 0.16$, 95% CI = [−.004, .33], $t(114) = 0.05$, $p = .054$ (see Figure 5; we also compared their absolute ratings, see SM).

Thus, we found that voice-influenced images were judged differently when starting from negative or positive faces and listening to inconsistent than consistent voices.

General Discussion

Across two experiments, we tested whether listening to a valence-inconsistent voice can alter face representations created by RC. In Experiment 1, we found that vocal information changed classification images of face representations formed from negative behavioral information. Participants who initially evaluated a target negatively updated face representations more positively after hearing a positive (vs. negative) voice, an effect consistent with explicit evaluations. In Experiment 2, we examined whether face representations created from a glimpse of a target's face would similarly shift when followed by valence-inconsistent voices. Images were judged as more positive when a negative face was paired with a positive voice, and slightly more negative when a positive face was paired with a negative voice, even though explicit evaluations from images producers did not differ significantly.

These findings extend prior work on impression formation, showing that impressions can be formed and updated cross-modally (Masi et al., 2022; Shen et al., 2020). Experiment 1 showed that self-reported negative behavior-based impressions could be updated by inconsistent voices (Masi et al., 2024). However, explicit judgments alone tell just one side of the story: The difference in judgment does not necessarily reveal the alteration of face representations. Crucially, using RC, here we filled such a gap demonstrating that not only explicit impressions—shaped by assumptions about the target's character (i.e., "who the person is")—but also face representations are updated by voices (i.e., "how the person—may—look"). This extends previous RC findings overlooking cross-modal interferences (Klein et al., 2021; Rougier & De Houwer, 2025). Whether these changes in face representations reflect persistent memory alterations or temporary shifts remains an open question. Future research should examine the stability of these modified representations over time (Cone et al., 2021b).

Experiment 2 demonstrates that face representations, even when formed without explicit instructions, can be influenced by voices. Participants were told to memorize faces before hearing voices and doing the RC task, where the resulting images indicated voice-induced modifications, though weaker than those observed in Experiment 1. Unlike Experiment 1, explicit ratings in Experiment 2 did not reveal differences in final impressions. This dissociation between RC and self-report may reflect differences in the intuitiveness of RC responses. In Experiment 1, an explicit evaluation at T1 preceded the RC task, potentially prompting participants to select noisy faces congruent with their reported impressions. The omission of this step in Experiment 2 may have reduced the intuitiveness of RC decisions, thereby weakening their alignment with explicit ratings. As Axt et al. (2023) argue, greater intuitiveness of findings increases the likelihood of convergence between RC and self-report. Together, these results underscore the unique contribution of RC for capturing changes in spontaneously formed representations beyond self-reports and highlight the value of multimethod approaches for understanding cross-modal impression formation.

This research has methodological limitations. First, participants were sampled with replacement across iterations for the images generation, and we did not track whether the same individuals contributed to both T1 and T2 subgroups—limiting our ability to statically model within-subject dependencies. Second, Experiment 2 used different base faces across conditions, which may introduce variability. Although this approach ensured consistency between the face seen at T1 and those used for the RC task, it may have reduced comparability across conditions. Future studies could test alternative designs using a common morphed face for RC. Finally, while we focused on valence, both voices and faces conveyed several unmeasured social traits. Future work should explore more in depth the mechanism, that is, whether trait-level changes—such as shifts in perceived dominance or trustworthiness—explain the influence of voice on visual representations, to better capture the dynamics of cross-modal impressions.

To conclude, this research extends knowledge of person perception by showing that inconsistent voices can update the evaluative content of face representations obtained via reverse correlation. These findings support the view that cross-modal processes dynamically integrate social information to shape person representations over time.

Acknowledgments

The authors thank Dr. Marco Petilli for his assistance with data analysis.

ORCID iDs

Matteo Masi https://orcid.org/0000-0001-6436-1663
Marco Brambilla https://orcid.org/0000-0002-4774-3309

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In Experiment 1, we assessed impression updating by comparing T1 and T2 classification images. In Experiment 2, only T2 images were available, as no RC task or judgment was collected at T1. This allowed us to test spontaneity of sequential cue integration but not updating, so we refer to Experiment 2 as assessing final representation differences (consistent vs. inconsistent voices) rather than impression updating.
2. We discuss the potential limitations of our exclusion criterion in the SM.
3. In the preregistration, we included time as a factor because of the sequential cue presentation, but since we do not measure impressions and representations at T1, we excluded it.
4. The preregistration cited "b = .08" due to a typo.
5. We measured surprise to explore whether expectancy violation explained updating; however, no association was found. Due to space constraints, these results are reported in SM.

References

Axt, J., Siemers, N., Discepola, M. N., Martinez, P., Xiao, Z., & Wehrli, E. (2023). The mind's "aye"? Investigating overlap in findings produced by reverse correlation versus self-report. *Journal of Experimental Social Psychology*, *107*, Article 104473. https://doi.org/10.1016/j.jesp.2023.104473

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Bülthoff, I., & Newell, F. N. (2017). Crossmodal priming of unfamiliar faces supports early interactions between voices and faces in person perception. *Visual Cognition*, *25*(4-6), 611–628. https://doi.org/10.1080/13506285.2017.1290729

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021a). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological and Personality Science*, *12*(5), 760–768. https://doi.org/10.1177/1948550620938616

Cone, J., Flaharty, K., & Ferguson, M. J. (2021b). The long-term effects of new evidence on implicit impressions of other people. *Psychological Science*, *32*(2), 173–188. https://doi.org/10.1177/0956797620963559

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology*, *56*, 131–199. https://doi.org/10.1016/bs.aesp.2017.03.001

Dotsch, R. (2023). *rcicr: Reverse correlation image classification toolbox* (R package version 1.0.1). https://github.com/rdotsch/rcicr

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Dotsch, R., Wigboldus, D. H., & Van Knippenberg, A. D. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology*, *43*(1), 116–125. https://doi.org/10.1002/ejsp.1928

Farley, S. (2025). Introduction to the special issue on innovations in vocal research: Insights from emotion, personality perception, eyewitness accuracy and persuasion. *Journal of Nonverbal Behavior*, *49*, 3–7. https://doi.org/10.1007/s10919-025-00483-2

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. *Advances in Experimental Social Psychology*, *61*, 237–287. https://doi.org/10.1016/bs.aesp.2019.09.005

Green, P., & MacLeod, C. J. (2016). Simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Klein, S. A., Hutchings, R. J., & Todd, A. R. (2021). Revising mental representations of faces based on new diagnostic information. *Cognition*, *217*, Article 104916. https://doi.org/10.1016/j.cognition.2021.104916

Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C., & Garrido, L. (2017). Similar representations of emotions across faces and voices. *Emotion*, *17*(6), 912–937. https://doi.org/10.1037/emo0000282

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lavan, N. (2023). How do we describe other people from voices and faces? *Cognition*, *230*, Article 105253. https://doi.org/10.1016/j.cognition.2022.105253

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means* (R package version 181-1). https://CRAN.R-project.org/package=emmeans

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Masi, M., Mattavelli, S., Fasoli, F., & Brambilla, M. (2022). Cross-modal impression updating: Dynamic impression updating from face to voice and the other way around. *British Journal of Social Psychology*, *61*(3), 808–825. https://doi.org/10.1111/bjso.12511

Masi, M., Mattavelli, S., Fasoli, F., & Brambilla, M. (2023). When faces and voices come together: Face width-to-height ratio and voice pitch contribute independently to social perception. *European Journal of Social Psychology*, *53*(2), 323–335. https://doi.org/10.1002/ejsp.2905

Masi, M., Mattavelli, S., Fasoli, F., & Brambilla, M. (2024). Multi-modal cues to change your mind: The intertwining of faces, voices, and behaviors in impression updating. *Personality and Social Psychology Bulletin*. Advance online publication. https://doi.org/10.1177/01461672241273209

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McAleer, P., & Belin, P. (2018). The perception of personality traits from voices. In S. Fruholz, & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 585–606). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.26

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLOS ONE*, *9*(3), Article e90779. https://doi.org/10.1371/journal.pone.0090779

McArthur, L. Z., & Baron, R. M. (1983). Toward and ecological theory of social perception. *Psychological Review*, *90*(3), 215–238. https://doi.org/10.1037/0033-295X.90.3.215

Mileva, M. (2025). Multimodal person evaluation: First impressions from faces, voices, and names. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspa0000454

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, *13*(2), 248–277. https://doi.org/10.1016/0010-0285(81)90010-4

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319–332. https://doi.org/10.1177/1745691614528519

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, *39*(4), 355–370. https://doi.org/10.1007/s10919-015-0214-8

Rougier, M., & De Houwer, J. (2025). Unconstraining evaluative conditioning research by using the reverse correlation task. *Social Psychological and Personality Science*, *16*(3), 291–299. https://doi.org/10.1177/19485506231217526

Schmitz, M. (2025). *schmitz* (R package version 0.1.3). https://github.com/matschmitz/schmitz

Schmitz, M., Rougier, M., & Yzerbyt, V. (2024). Introducing the brief reverse correlation: An improved tool to assess visual representations. *European Journal of Social Psychology*, *54*(7), 1446–1461. https://doi.org/10.1002/ejsp.3100

Shen, X., Maiolatesi, A. J., Mann, T. C., & Ferguson, M. J. (2023). The relation between updated implicit evaluations and the trust game. *Social Psychological and Personality Science*, *14*(8), 911–921. https://doi.org/10.1177/19485506221141661

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face?: Updating face-based implicit impressions using diagnostic behavioral information. *Journal of Experimental Social Psychology*, *86*, Article 103888. https://doi.org/10.1016/j.jesp.2019.103888

Skowronski, J. J., Carlston, D. E., & Hartnett, J. (2008). Spontaneous impressions derived from observations of behavior: What a long, strange trip it's been (and it's not over yet). In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 313–333). Guilford Press.

Sutherland, C. A., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*, *113*(4), 1056–1078. https://doi.org/10.1111/bjop.12583

Todorov, A., Oh, D., Uddenberg, S., & Albohn, D. N. (2025). Face evaluation: Findings, methods, and challenges. *Annals of the New York Academy of Sciences*, *1545*(1), 28–37. https://doi.org/10.1111/nyas.15293

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Vives, M. L., Frances, C., & Baus, C. (2023). Automating weighing of faces and voices based on cue saliency in trustworthiness impressions. *Scientific Reports*, *13*(1), Article 20037. https://doi.org/10.1038/s41598-023-45471-y

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

Author Biographies

Matteo Masi is a postdoctoral researcher at the University of Milano-Bicocca, where he conducts research on social cognition, with a focus on first impressions based on faces and voices. He earned his PhD in social psychology from the University of Milano-Bicocca and the University of Surrey.

Marco Brambilla is full professor and director of the social perception and cognition Lab at the University of Milano-Bicocca. He investigates how people form impressions of others based on facial, vocal, and behavioral cues as well as the processes through which these impressions dynamically change over time. He also examines how individuals perceive and evaluate one another across group boundaries and how stereotypes and prejudice influence impression formation and social behavior.