Received: 26 May 2021 | Accepted: 16 November 2021

DOI: 10.1111/bjso.12511

ORIGINAL ARTICLE

the british psychological society
promoting excellence in psychology

Cross-modal impression updating: Dynamic impression updating from face to voice and the other way around

Matteo Masi[1,2] | **Simone Mattavelli**[1] | **Fabio Fasoli**[2] | **Marco Brambilla**[1]

[1]University of Milano-Bicocca, Milano, Italy

[2]University of Surrey, Guildford, UK

Correspondence
Matteo Masi, Department of Psychology, University of Milano-Bicocca, Piazza dell' Ateneo Nuovo, 1, 20126 – Milano, Italy.
Email: m.masi8@campus.unimib.it

Abstract

Research has shown that faces and voices shape impression formation. Most studies have examined either the impact of faces and voices in isolation or the relative contribution of each source when presented simultaneously. However, only a few studies have questioned whether and how impressions formed via one source can be updated due to incremental information gathered from the alternative source. Yet, cross-modal impression updating is key to shed light on person perception. Thus, we tested whether positive and negative face- and voice-based impressions could be updated by inconsistent cross-modal information. In Experiment 1 (N = 130), we tested whether face-based impressions could be updated by (in)consistent voices. In Experiment 2 (N = 262), we compared face-to-voice and voice-to-face impression updating. In Experiment 3 (N = 242), we favoured a more direct comparison of the two types of stimuli (i.e., the co-occurrence of both cue types when the new information is revealed). Results showed that voices have the greatest updating impact and that the updating effect of faces was halved when voices co-occurred for a second time. We discussed these results as evidence of the dynamical evolution of cross-modal impressions.

KEYWORDS
cross-modal integration, face perception, impression updating, voice perception

INTRODUCTION

Imagine yourself sitting in the bus and scrolling down your favourite social network's homepage on your smartphone. You get a notification showing the latest picture of Paul, a good friend of yours, together with a new acquaintance named Mark. In his profile picture, Mark does not look like a trustworthy person, and you are not keen to meet him. At the next stop, Paul and Mark get on the bus and run into you. Paul introduces you to Mark, who says, “Hi, nice to meet you!”. His voice sounds quite nice so that your initial impression suddenly changes for the better. In everyday life, both the face and the voice of our interaction partners shape impressions. Here, we tested whether vocal and facial cues dynamically interact in influencing our impressions of others. Specifically, we examined the possibility that voices may update face-based impressions and vice versa. Similarly, we tested whether one specific cue updates cross-modal impressions to a greater extent than the other.

Impression formation from faces and voices

Non-behavioural cues are information used to understand others’ intentions and personalities and discriminate whom to avoid or approach (Ambady et al., 2000; Brooks & Freeman, 2018; Zebrowitz & Montepare, 2008). People spontaneously judge others from their faces (Todorov et al., 2015; Zebrowitz, 2017) and few milliseconds are sufficient to form a face-based impression about the person (Olivola & Todorov, 2010; Willis & Todorov, 2006). From a glance, we can nimbly infer a person's age, gender, ethnicity, feelings, political affiliation, or sexual orientation (e.g., Rule & Sutherland, 2017; Rule et al., 2015), and we can detect other's good or bad dispositions (Todorov et al., 2009, 2015). Moreover, face-based impressions can predict a great wealth of social outcomes, such as political voting (Ballew & Todorov, 2007; Olivola & Todorov, 2010) or court decisions (Wilson & Rule, 2015; see Todorov et al., 2015 for a review).

People also form impressions from voices (Aronovitch, 1976; McAller & Belin, 2018). Few spoken words or syllables are sufficient to form impressions about speakers’ physical characteristics, gender, age, emotions (e.g., Koolagudi & Rao, 2012; Pisanski et al., 2016), sexual orientation, or personality traits (McAller et al., 2014; Sulpizio et al., 2015). Like impressions shaped by facial cues, voice-based impressions can influence voting behaviour (Klofstad et al., 2012; Tigue et al., 2012) and court decisions (Scherer, 1979; see McAller & Belin, 2018 for a review). Thus, both faces and voices are rich sources of information that ultimately shape our judgements and behaviours towards others.

Most studies have investigated face- and voice-based impressions separately (McAller & Belin, 2018; Todorov et al., 2015). However, our social and perceptual life is rarely unimodal. Efficiently combining distinct perceptual cues is functional to categorize people into social categories (e.g., ethnicity, Rakić et al., 2011, 2020) and infer their affective states (Campanella & Belin, 2007; De Gelder & Berteleson, 2003). Thus, a more accurate way to establish the weight of both faces and voices in shaping impressions may require that the contribution of each source be studied net of the other one. Indeed, recent research has investigated the relative contribution of the face–voice simultaneous perception on impression formation showing that face–voice combination influenced group-based impressions. For instance, it has been shown that ethnic cues conveyed by both facial (i.e., ethnic phenotypes) and the vocal (i.e., accent) features jointly influence intergroup impressions, with people being more affected by others’ voices (Hansen et al., 2017a, 2017b). Face–voice combination also influences interpersonal impressions (Mileva et al., 2018; Rezlescu et al., 2015; Zuckerman & Sinicopri, 2011). Moreover, the relative contribution of either the face or the voice in shaping interpersonal impressions depends on the dispositional trait under investigation. While someone's face matters more when it comes to attributing trustworthiness (Mileva et al., 2018), voices have higher informative power when it comes to evaluating others' dominance (Rezlescu et al., 2015). However, their contribution to global impressions has remained untested so far.

Importantly, the face and the voice of another person are often processed sequentially and can dynamically influence each other, modifying our impressions about that person. For instance, we can form an impression from a person's picture and then update it when listening to their voice on the phone, or vice versa. In such cases, impressions from faces and voices are not the mere result of two information's simultaneous integration. Rather, they originate from one source and are subsequently updated based on the alternative source.

Impression updating with faces and voices

Social stimuli's functional role is not to create ever-lasting impressions that plaster up the perceiver but to prepare the most adaptive reaction towards the social interaction partner (McArthur & Baron, 1983). Under this view, impressions would be updated if the new information is perceived as more revelatory of the interlocutor's real intentions than the previous one (Brambilla et al., 2019). Indeed, impressions can evolve dynamically (for a review, Freeman et al., 2020). Studies have tested how impressions originated from faces could be updated after exposure to inconsistent behavioural information (i.e., McConnel et al., 2008; Shen & Ferguson, 2021; Shen et al., 2020). However, empirical studies on impression updating considering only non-behavioural information gathered from distinct sources are lacking. One exception can be found in Hansen et al. (2018). The authors examined the joint impact of faces and voices on the target's ethnicity within the context of hiring decisions. Job candidates who sounded native (i.e., German-accented individual) but later looked foreign (i.e., Turkish-looking individual) were ultimately evaluated based on the face, whereas the opposite happened when the presentation order of the cues was reversed, with the voice influencing the final evaluation of the candidate. Yet, these findings are confined to ingroup–outgroup dynamics triggered by specific (facial and vocal) features. Moreover, Hansen and colleagues measured the evaluation of the target within a context (i.e., hiring decision) in which ethnicity plays a key role (Shore et al., 2009). Thus, it remains unknown how faces and voices combine in shaping interpersonal impressions that are not constrained by specific ingroup–outgroup stereotypes.

Furthermore, in Hansen et al. (2018) participants were first exposed to the first information (i.e., face or voice) and reported their impression. Next, half a second later, the participants were exposed to the identical initial cue and to the cross-modal new information before reporting a second evaluation. Although sequential, this procedure implied that the second cue was never revealed in isolation, which prevented one from assessing its unique contribution to impression updating. Thus, no prior work has measured the individual contribution of cross-modal cues presented in isolation in updating initial impressions. Filling these gaps, our research considered the interplay between faces and voices in forming interpersonal impressions by further aiming to establish which one drives the strongest cross-modal impression updating.

The present research

The present work extended prior research by testing the face–voice interaction in impression updating. Across three studies, we presented individuals' faces and voices (manipulated in reflectance/shape and pitch, respectively, to reflect variations in valence) sequentially and individually, and measured the global impression upon the presentation of each cue. In Experiment 1, we started from face-based impressions of individuals and tested whether those impressions could be updated by their own voices processed at a later stage and manipulated to be either consistent or inconsistent in terms of valence. Experiment 2 was a direct comparison between face-to-voice and voice-to-face impression updating. Testing the role of individuals' faces and voices presented sequentially might help to determine which cue leads to stronger impression updating through the alternative channel. Finally, in Experiment 3, we compared impression updating with cross-modal cues presented sequentially and individually with a scenario in which cross-modal cues co-occurred with the first cue. Comparing those two scenarios allowed us to investigate whether the presentation modality of the second stage (isolation vs. co-occurrence) would influence impression updating.

All the experiments were conducted online with a direct link to Inquisit Web 4.0.4 for Experiment 1 and Qualtrics for Experiments 2–3. Participants were volunteers recruited through social networks. None reported being affected by hearing or vision disparities. Significant interactions were inspected with simple effects analysis.[1] The studies have been approved by the local university ethics committee. All the materials, data, and analysis code are available at https://osf.io/up62v/.

EXPERIMENT 1

Experiment 1 explored whether face-based impressions could be updated after being exposed to the voice of the target individual. Voices could be either consistent or inconsistent with the information provided by the face. We considered this study as explorative and, therefore, we did not set *a priori* expectations on whether and how voices would update positive and negative face-based impressions. We employed a 2 (face valence: positive vs. negative) × 2 (face–voice consistency: consistent vs. inconsistent) × 2 (time: time 1 vs. time 2) mixed design, with the second factor manipulated between participants.

Method

Participants and procedure

One hundred and thirty-six Italian participants completed the experiment. Six participants failed a memory check and were therefore excluded. The final sample consisted of 130 participants ($n \approx 65$ per between-subjects cell, $M_{age} = 23.27$, $SD_{age} = 6.66$, 76 females). A sensitivity power analysis (MorePower v6.0.4, Campbell & Thompson, 2012) suggested that $N = 130$ participants could detect an effect as small as $f = .25$ ($\eta_p^2 = .058$) with power = 80% (at alpha = .05) in a 2 × 2 × 2 mixed design with one between-participant factor.

Participants were invited to take part in an online experiment about impression formation. They were first instructed to put earphones on, check the functioning of their equipment, and set the volume of their device at a reasonable level. The experiment consisted of two consecutive blocks. In the first block, participants were presented with a positive and a negative face belonging to two individuals and then were asked to provide a judgement on a 7-point scale (−3 = negative; +3 = positive, evaluation at time 1). Afterwards, participants listened to each person's voice, either consistent or inconsistent with the face valence. Thus, they provided a second judgement (evaluation at time 2). Finally, they reported their demographics (age, gender) before being thanked and debriefed.

Stimuli

Faces

Two pairs of positive and negative faces were extracted from a validated database (Todorov et al., 2013) in which digital faces varied with a linear increase on a continuum of trustworthiness, ranging from −3*SD* to +3*SD* in steps of 1*SD*. For both the types of faces (i.e., positive and negative), we used the mid-level (−/+2*SD*) manipulations. We employed stimuli varying in facial trustworthiness, as studies demonstrated that trustworthiness is a key driver of impressions (Brambilla et al., 2021; Riva et al., 2016), including those

[1]In the simple effects analyses, no factor presented more than two levels. Therefore, a single *t*-test was performed within each level of the other factors (or factorial combinations). Indeed, no multiplicity issue was created, eliminating the necessity for the *p*-value adjustment.

based on facial cues (Oosterhof & Todorov, 2008; Todorov et al., 2008). Furthermore, as these were digital identities, they were free from any confounding variable that may influence the ultimate evaluation (e.g., hair colour and style, skin imperfections), an operation that is hardly accomplished with real photographs.

Voices

Two pairs of positive and negative voices were created. First, 16 Italian male students (M_{age} = 25.50, SD_{age} = 4.28) were recorded while uttering six sentences neutral in semantic content, $t(59) < 1.85$, $p > .07$ (see Appendix S1). Speakers were recorded alone and instructed to speak in a neutral way. Next, 80 students (M_{age} = 23.51, SD_{age} = 7.97, 35 Italians, 45 British)[2] evaluated each speaker on a 7-point scale (−3 = negative; +3 = positive). Two speakers whose ratings did not differ from the scale midpoint were selected: M_a = 0.05, SD_a = 1.03; M_b = 0.18, SD_b = 1.06), $t_a(77) = 0.57$, $p = .57$, $d = .06$, 95% CI [−0.16, 0.29]; $t_b(78) = 1.09$, $p = .27$, $d = .12$, 95% CI [−0.10, 0.35]. No difference emerged between these two, $t(76) = 0.73$, $p = .46$, $d = .08$, 95% CI [−0.31, 0.14]. Hence, the two speakers' voices were perceived as neutral.

Second, we created a negative and a positive variation of each selected neutral voice. The male voice's mean fundamental frequency *f*0 (hereafter, pitch) shares a positive relationship with valence (McAller et al., 2014). Therefore, we assumed that a regular increase or decrease of 5Hz, 7 Hz, or 9 Hz (Audacity Team, 2019) applied to the original speakers' mean pitch value (M_a = 98.56 Hz, SD_a = 1.68 Hz; M_b = 102.5 Hz, SD_b = 3.80 Hz) would create positive and negative voices, respectively. To confirm our assumptions, 57 participants (M_{age} = 25.19, SD_{age} = 8.14, 35 females) evaluated the digitally manipulated voices on a 7-point scale (−3 = negative; +3 = positive). For both speakers, the low-pitched voices (9 Hz lower than the original pitch, M_A = −0.40, SD_A = 1.61; M_B = −0.42, SD_B = 1.36) were perceived as more negative than the high-pitched voices (5 Hz higher than the original pitch, M_A = 0.75, SD_A = 1.33; M_B = 0.56, SD_B = 1.54), $t_a(56) = −4.56$, $p < .001$, $d = .60$, 95% CI [0.32, 0.89]; $t_b(56) = −4.15$, $p < .001$, $d = .55$, 95% CI [0.27, 0.83]. Also, the two negative voices did not differ from each other, $t(56) = 0.07$, $p = .94$, $d = .01$, 95% CI [−0.25, −0.27], as did not the two positive voices, $t(56) = 0.85$, $p = .40$, $d = .11$, 95% CI [−0.15, 0.38].

This procedure allowed us to obtain voices that were neutral in content and intonation but varied in valence due to pitch manipulation.

Face presentation

In the first block, participants were informed that they would 'meet' two people, Marco and Luca, whose faces would be presented in two alternated times. They were asked to memorize the name associated with each face, displayed above each visual stimulus, as this information linked the first and the second information to the same identity. These faces were one positive and one negative presented in a counterbalanced order across participants. Each face remained on screen for approximately 12 s. Right after, a memory check item assessed whether participants remembered the name associated with each face. Then, we measured participants' impressions as illustrated in the procedure.[3]

Voice presentation

In the second block, participants were presented with the (in)consistent voices. Each identity was introduced by the name, one at a time and in a counterbalanced order. Meanwhile, a random sequence of

[2]We recruited British participants who did not speak Italian to further control that the evaluation was not affected by the sentences' semantic content. Ratings did not differ across the multilinguistic samples, $F(1, 78) = 2.23$, $p = .14$, $\eta_p^2 = .03$, 95% CI [0, 0.11].

[3]For exploratory purpose, we also administered two Implicit Association Tests (IATs, Greenwald et al., 1998). The order of administration of the IAT and the explicit measure was alternated between participants. Since the IAT was not affected by our manipulation, results are not further discussed but reported in Appendix S1.

three audio recordings uttered by the same voice for each identity was played back with an interstimulus pause of 1000 ms. The speaker's voices were presented for a total of 12 s each. Then, participants answered the name memory check item, before reporting their second evaluation.

Results

Impression formation

In line with the idea that people form impressions based on facial features, positive faces (M = 0.85, SD = 1.10) were judged more positively than negative faces (M = −0.25, SD = 1.28), $t(129) = 6.55$, $p < .001$, $d = .57$, 95% CI [0.22, 0.92].

Impression updating

Updating of initial impressions consisted in the difference between ratings provided by participants at time 1 versus time 2. For the sake of simplicity, we calculated a unique differential score from the two original ratings, rather than entering the two scores as within-subject outcomes: we subtracted time 2 scores from time 1 scores when the first information was positive, and time 1 scores from time 2 scores when negative. Thus, the more positive the score, the greater the impression updating was. We performed a 2 (face valence: positive vs. negative) × 2 (face–voice consistency: consistent vs. inconsistent) mixed analysis of variance (ANOVA) (see Table 1 and Figure 1). We found a significant main effect of consistency, $F(1, 128) = 18.39$, $p < .001$, $d = .76$, 95% CI [0.40, 1.12], suggesting that the inconsistent voices updated the initial impression more than the consistent ones. As the calculated updating score reflected the differences between the two evaluations provided after processing faces and then voices, a score greater than zero would mean that the second piece of information (i.e., the voice) significantly updated the first impression. Contrasting either score to zero in one-sample t-tests revealed a large updating due to inconsistent voices, $t(127) = 5.99$, $p < .001$, $d = 1.06$, 95% CI [0.69, 1.43], but no updating when the voice was consistent, $t(131) = −0.83$, $p = .41$, $d = .14$, 95% CI [−0.49, 0.20]. Neither a significant main effect of face valence nor a significant interaction between the two factors emerged, $F(1, 128) = 2.19$, $p = .14$, $d = .13$, 95% CI [−0.04, 0.30] and $F(1, 128) = 0.64$, $p = .42$, $\eta_p^2 = .005$, 95% CI [0, 0.06], respectively. Thus, impression updating from faces to inconsistent voices emerged, regardless of whether the face was either positive or negative.

Discussion

Experiment 1 provided initial evidence in favour of the updating of face-based impressions through inconsistent voices, which emerged irrespective of the valence of the first information. In other words, participants updated their initial (face-based) impression of the target individual when such voices were inconsistent with the valence carried by the face.

EXPERIMENT 2

Experiment 2 aimed to replicate and extend Experiment 1 by further considering how people update voice-based impressions when exposed to targets' faces at a later stage. We employed a 2 (first information type: face vs. voice) × 2 (first information valence: positive vs. negative) × 2 (face–voice consistency: consistent vs. inconsistent) × 2 (time: time 1 vs. time 2) mixed design, with the first and the third factors manipulated between participants. Thus, participants were presented with either the face or the voice of the target individual at time 1 and with the (in)consistent alternative cue at time 2. This design allowed us to test whether the order in which the two pieces of information were delivered (i.e.,

TABLE 1 Means and standard deviations (in parentheses) of impressions formed on faces (positive vs. negative) at two alternated times (time 1 vs. time 2) as a function of consistency of the additional voice (consistent vs. inconsistent). Original scores range from −3 (negative) to +3 (positive). The column 'Score' represents the difference between time 1 and time 2

	Face-based impression					
	Positive			Negative		
Consistency	Time 1	Time 2	Score	Time 1	Time 2	Score
Consistent	0.71 (1.16)	0.97 (1.02)	−0.26 (1.34)	−0.24 (1–20)	−0.18 (1.19)	0.06 (1.39)
Inconsistent	0.98 (1.03)	0.20 (1.24)	0.78 (1.60)	−0.25 (1.37)	0.62 (1.20)	0.87 (1.54)

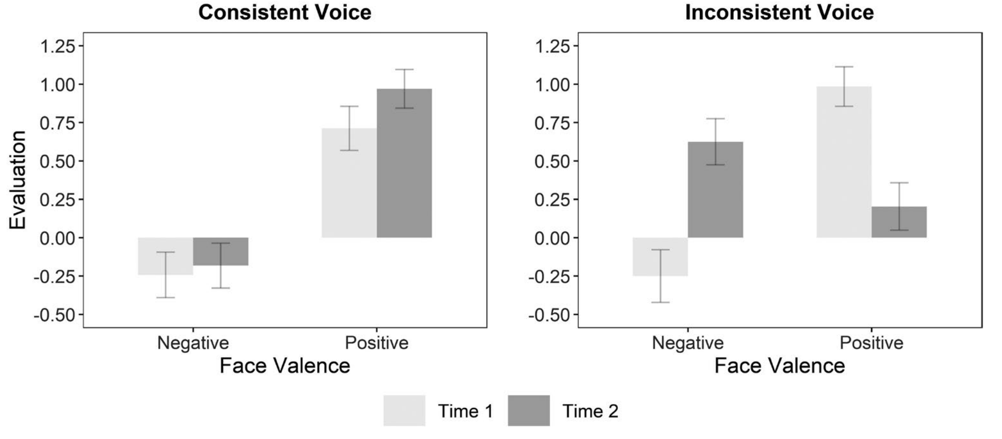

FIGURE 1 Bar graphs of Experiment 1 depicting evaluations given at time 1 (light grey bar) and time 2 (dark grey bar) as a function of face valence (negative vs. positive) and face–voice consistency (consistent vs. inconsistent). Evaluations are expressed on the original scale ranging from −3 (negative) to +3 (positive)

face-to-voice vs. voice-to-face) affected impression updating. Showing that the type of the second information impacts upon impression updating would speak in favour of the superiority of either cue to update impressions. Alternatively, the two first impressions could be updated identically, suggesting that the two cues similarly account for impression updating.

Moreover, readers might argue that the effect found in Experiment 1 could confine to the set of stimuli employed. Considering that voices were created through a series of time-consuming pre-tests (see Experiment 1), we decided to test whether the effect observed in Experiment 1 could generalize over new faces. Hence, two new faces, one depicting a trustworthy and one an untrustworthy man, were extracted from the validated database employed in Experiment 1 (Todorov et al., 2013).

Method

Participants and procedure

Two hundred and sixty-two Italian participants ($n \approx 65$ per between-subjects cell, $M_{age} = 30.59$, $SD_{age} = 13.75$, 165 females) completed the experiment.[4] A sensitivity power analysis (MorePower v6.0.4)

[4]An error in programming the vocal identities' memory test mismatched the sentences pronounced by each individual in the main task with the sentences pronounced in the test session, making it impossible to quantify the precise number of people who failed it because of this error or because of their lack of attention. We decided to keep all participants in this analysis and report the analysis without them ($N = 164$) in Supplementary Materials.

returned that, with N = 262 participants, this experiment would detect an effect as little as f = .17 (η_p^2 = .028) with power = 80% (at alpha = .05) in a 2 × 2 × 2 × 2 mixed design with two between-subject factors. Instructions and procedure were identical to Experiment 1, except for the two alternated orders of presentation of faces and voices. Judgements were assessed again on a 7-point scale (−3 = negative, +3 = positive).

Results

Impression formation

Impressions reported at time 1 were submitted to a 2 (first information valence: positive vs. negative) × 2 (first information type: face vs. voice) mixed ANOVA. We found a significant main effect of valence, $F(1, 260) = 34.00$, $p < .001$, $d = .36$, 95% CI [0.24, 0.49]. Positive cues ($M = 0.85$, $SD = 1.32$) elicited more positive impressions than negative cues ($M = 0.24$, $SD = 1.41$). Also significant was the effect of information type, $F(1, 260) = 6.45$, $p = .01$, $d = .31$, 95% CI [0.07, 0.56], suggesting that voice-based impressions ($M = 0.71$, $SD = 1.33$) were overall more positive than face-based impressions ($M = 0.38$, $SD = 1.44$). The interaction between the two factors was also significant, $F(1, 260) = 11.54$, $p < .001$, $\eta_p^2 = .04$, 95% CI [0.01, 0.10]. Negative faces elicited more negative impressions ($M = -0.10$, $SD = 1.39$) than negative voices ($M = 0.58$, $SD = 1.35$), $t(260) = -4.11$, $p < .001$, $d = .37$, 95% CI [0.19, 0.55], but such a difference did not emerge for positive cues ($M_{\text{Faces}} = 0.87$, $SD = 1.34$ vs. $M_{\text{Voices}} = 0.84$, $SD = 1.31$), $t(260) = 0.17$, $p = .86$, $d = .02$, 95% CI [−0.16, 0.19].

Impression updating

The impression updating score was computed as in Experiment 1. We inspected the updating of first impressions due to the cross-modal cue's addition to face- or voice-based first impressions. Hence, a 2 (first information valence: positive vs. negative) × 2 (first information type: face vs. voice) × 2 (face–voice consistency: consistent vs. inconsistent) mixed ANOVA was performed (see Table 2 and Figure 2). Replicating Experiment 1, a main effect of consistency revealed that inconsistent cues updated initial impressions more than consistent cues, $F(1, 260) = 33.95$, $p < .001$, $d = .73$, 95% CI [0.47, 0.98]. A series of one-sample t-tests against zero showed that both inconsistent cues, $t(257) = 6.06$, $p < .001$, $d = .76$, 95% CI [0.50, 1.01], and consistent cues updated the first impressions, $t(265) = -2.07$, $p = .04$, $d = .25$, 95% CI [0.01, 0.50]. Furthermore, a significant effect of first information type, $F(1, 260) = 7.57$, $p = .006$, $d = .31$, 95% CI [0.07, 0.56], indicated that voices updated face-based impressions more than vice versa. One-sample t-tests against zero showed that voices significantly updated face-based impressions, $t(259) = 4.17$, $p < .001$, $d = .52$, 95% CI [0.27, 0.77], while faces did not update voice-based impressions, $t(263) = 0.41$, $p = .68$, $d = .05$, 95% CI [−0.19, 0.29]. Finally, the interaction between first information type and valence was significant, $F(1, 260) = 9.00$, $p = .003$, $\eta_p^2 = .04$, 95% CI [0.01, 0.10]. Negative face-based impressions were updated by voices more than voice-based impressions were updated by faces, $t(510) = 4.06$, $p < .001$, $d = .36$, 95% CI [18, 0.53], while no difference occurred between positive first impressions, $t(510) = 0.04$, $p = .96$, $d = .004$, 95% CI [−0.17, 0.18]. The t-tests against zero showed that only negative face-based impressions were updated by the additional cue, $t(129) = 4.33$, $p < .001$, $d = .76$, 95% CI [0.40, 1.12], while none of the others was: positive face-based impressions: $t(129) = 1.51$, $p = .13$, $d = .27$, 95% CI [−0.08, 0.61], negative voice-based impressions: $t(131) = -0.91$, $p = .36$, $d = .16$, 95% CI [−0.18, 50], positive voice-based impressions: $t(131) = 1.53$, $p = .13$, $d = .27$, 95% CI [−0.08, 0.61]. No other significant main effect or interaction was found (ps > .29).

TABLE 2 Means and standard deviations (in parentheses) of face- and voice-based impressions formed on two types of cues (positive vs. negative) at two alternated times (time 1 vs. time 2) as a function of consistency of the additional cross-modal information (consistent vs. inconsistent). Original scores range from −3 (negative) to +3 (positive). The column 'Score' represents the difference between time 1 and time 2

	Positive			Negative		
Consistency	**Time 1**	**Time 2**	**Score**	**Time 1**	**Time 2**	**Score**
Face-based impression						
Consistent	0.95 (1.27)	1.12 (1.26)	−0.17 (1.43)	−0.08 (1.34)	0.06 (1.38)	0.14 (1.53)
Inconsistent	0.78 (1.41)	0.18 (1.49)	0.60 (1.73)	−0.12 (1.44)	1.08 (1.55)	1.20 (1.83)
Voice-based impression						
Consistent	0.98 (1.31)	1.09 (1.37)	−0.10 (1.26)	0.56 (1.31)	0.01 (1.33)	−0.54 (1.14)
Inconsistent	0.69 (1.30)	0.17 (1.38)	0.52 (1.62)	0.61 (1.40)	0.94 (1.21)	0.33 (1.75)

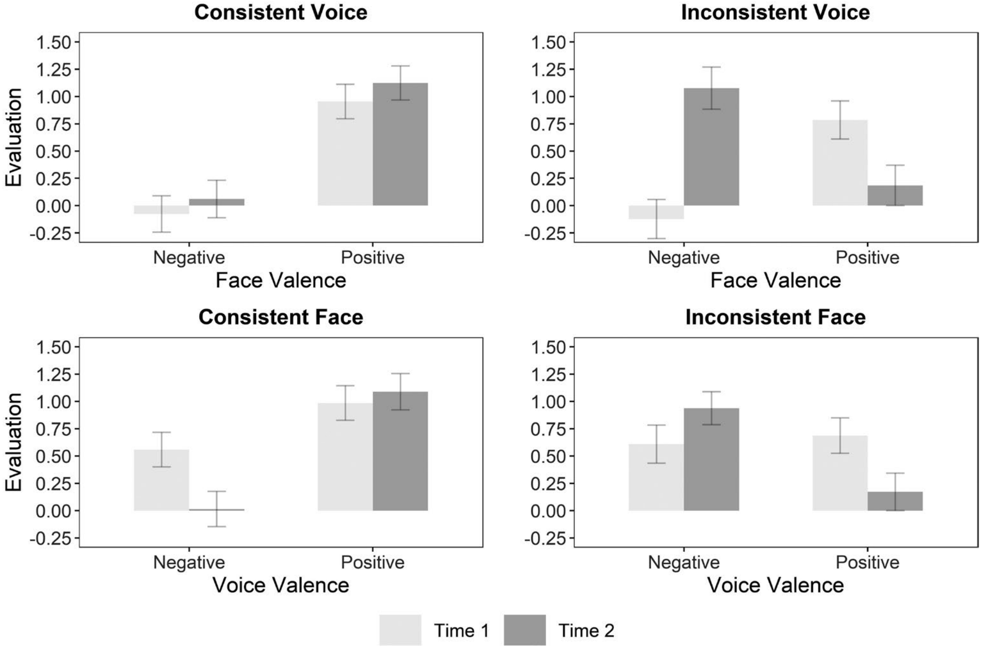

FIGURE 2 Bar graphs of Experiment 2 depicting evaluations given at time 1 (light grey bar) and time 2 (dark grey bar) as a function of first information valence (negative vs. positive) and face–voice consistency (consistent vs. inconsistent). The upper panels refer to face-to-voice impression updating, meaning that first impressions formed on faces were then updated by voices. The lower panels refer to voice-to-face impression updating, meaning that first impressions formed on voices were then updated by faces. Evaluations are expressed on the original scale ranging from −3 (negative) to +3 (positive)

Discussion

Experiment 2 showed that cross-modal cues could update both face- and voice-based impressions. Importantly, we found a stronger updating of face- (vs. voice-) based impressions via voices (vs. faces). Moreover, this effect was qualified by valence, such that voices updated negative, but not positive, face-based impressions of the individuals. Nonetheless, impression formation registered an asymmetry issue: when participants formed an impression based on target's negative face, such an impression was more extreme than the impression based on target's negative voice.

This asymmetry may have increased the likelihood of finding a difference between the two types of information because less extreme impressions could be more difficult to update due to the less room available (Hogarth & Einhorn, 1992). To account for such a possibility, we designed a third and final experiment with cross-modal stimuli more carefully balanced in valence. In Experiment 3, we focused only on inconsistent information delivered at time 2 and tested whether impression updating was somehow qualified by its presentation modality (i.e., in isolation vs. co-occurring with the initial cue).

EXPERIMENT 3

Cross-modal cues can update our impressions of a target person generated from both their faces and voices. Nevertheless, it has not been tested whether the way in which the second information is delivered impacts impression updating. So far, the effect was tested with cues presented in isolation once an impression has formed based on faces and voices. However, in everyday life, the additional cue (i.e., one's face or voice) can be encountered either in isolation or together with the initial one. By focusing only on the condition in which information is inconsistent in valence, here we compared these two scenarios and verified whether the first information re-presentation at time 2 would alter the effect of the inconsistent cue (for a similar although not identical procedure, see Hansen et al., 2018). Hence, in this study, we employed a 2 (first information type: face vs. voice) × 2 (first information valence: positive vs. negative) × 2 (time 2 presentation: in isolation vs. simultaneous) × 2 (time: time 1 vs. time 2) mixed design, with the first and the third factors varying between participants.

This design would unveil whether the superior updating effect of voices on face-based impressions found in Experiment 2 could be moderated by how participants encounter such information. Indeed, re-presenting the first information (e.g., negative face) along with the new inconsistent information (e.g., positive voice) would call for a more overt comparison between the two types of information, which could either amplify or decrease the updating of cross-modal impressions. Importantly, such a co-occurrence of the two stimuli at time 2 is functional to test an alternative impression development route and does not overlap with the procedure employed by previous studies considering how two cues simultaneously interact in shaping first impressions (Mileva et al., 2018).

Participants and procedure

Two hundred and seventy-six Italian participants volunteered to take part in the experiment. However, 34 participants failed the memory check test, leaving a sample of 242 (M_{age} = 25.35, SD_{age} = 12.64, 180 females) participants. A sensitivity power analysis (MorePower v6.0.4) returned that with N = 242 participants we could detect an effect as little as f = .18 (η_p^2 = .032) with power = 80% (at alpha = .05) in a 2 × 2 × 2 × 2 mixed design with two between-subjects factors.

The procedure largely mirrored that adopted in Experiment 2, except for the following changes. First, the individuals' face or voice introduced at time 2 was always of opposite valence compared to the first information. Second, for half of the participants, the two pieces of information were presented in isolation, while for the other half the second information was presented in co-occurrence with the first one at time 2. In this new condition, participants were presented with either a face or a voice at time 1 and with both the face and the voice at time 2 for a total of 12 s.

Since negative faces and voices' impressions were unbalanced in Experiment 2, we performed a pitch manipulation of −15Hz, instead of −9Hz, to increase the negativity of voice-based impressions. Faces were the same as in Experiment 2. Judgements were assessed again on a 7-point scale (−3 = negative, +3 = positive).

Results

Impression formation

Impression formation was measured in a 2 (first information valence: positive vs. negative) × 2 (first information type: face vs. voice) mixed ANOVA. The significant main effect of valence of the first information distinguished the impressions generated by positive ($M = 1.17$, $SD = 1.15$) and negative cues ($M = -0.50$, $SD = 1.35$), $F(1, 240) = 203.42$, $p < .001$, $d = .92$, 95% CI [0.77, 1.07]. The main effect of first information type was not significant, $F(1, 240) = .23$, $p = .63$, $d = .06$, 95% CI [−0.19, 0.31], but its interaction with valence was, $F(1, 240) = 4.55$, $p = .03$, $\eta_p^2 = .02$, 95% CI [0, 0.07]. Negative faces ($M = -0.41$, $SD = 1.26$) and voices ($M = -0.60$, $SD = 1.43$) were not different in their extremity, $t(477) = 1.24$, $p = .21$, $d = .11$, 95% CI [−0.07, 0.29], while positive faces ($M = 1.02$, $SD = 1.23$) tended to be perceived slightly less positively than voices ($M = 1.33$, $SD = 1.03$), but without overcoming the significance threshold in a meaningful way, $t(477) = -1.89$, $p = .06$, $d = .17$, 95% CI [−0.01, 0.35].

Impression updating

We analysed impression updating in a 2 (first information valence: positive vs. negative) x 2 (first information type: face vs. voice) x 2 (time 2 presentation: in isolation vs. simultaneous) mixed ANOVA (see Table 3 and Figure 3). The significant main effect of time 2 presentation showed that the isolated presentation of the two cues facilitated updating compared to when they co-occurred, $F(1, 238) = 4.40$, $p = .04$, $d = .28$, 95% CI [0.02, 0.53]. The significant main effect of first information type revealed that voices led to a greater updating of face-based impressions than the other way around, $F(1, 238) = 8.42$, $p = .004$, $d = .37$, 95% CI [0.12, 0.63]. There was a non-significant effect of valence, $F(1, 238) = 3.22$, $p = .07$, $d = .12$, 95% CI [−0.01, 0.25]. The interaction between the type of first information and time 2 presentation was significant, $F(1, 238) = 3.75$, $p = .05$, $\eta_p^2 = .01$, 95% CI [0, 0.06]. Voices (vs. faces) updated face-based (vs. voice-based) impressions more when the two cues were presented simultaneously at time 2, $t(238) = 3.43$, $p < .001$, $d = .44$, 95% CI [0.19, 0.70], but they did not differ when the cue was presented in isolation, $t(238) = 0.65$, $p = .51$, $d = .08$, 95% CI [−0.17, 0.34]. No other significant main effect or interaction was found (ps > .22).

Discussion

Overall, Experiment 3 confirmed that individuals' voices updated impressions more than faces did. The significant interaction with time 2 presentation mode revealed that this was true when targets' voices were presented along with faces. Moreover, such an effect was not affected by valence. Importantly, Experiment 3 presented voices and faces that were balanced in terms of baseline valence, which led to comparable first impressions. Even though the statistical test informed that the findings of Experiment 2 were not directly replicated when stimuli were presented in isolation, a closer look at the descriptive statistics suggested a similar pattern. To test whether inconsistent voices' greater updating effect holds across experiments, we performed a joint analysis of Experiment 2 and 3 data.

JOINT ANALYSIS OF EXPERIMENTS 2 AND 3

Although the two experiments yielded different results, we performed a joint analysis of their data ($N = 248$, face-to-voice updating = 128, voice-to-face updating = 120) to inspect with greater statistical power whether voices presented in isolation updated impressions more than faces. We performed a 2

TABLE 3 Means and standard deviations (in parentheses) of face- and voice-based impressions formed on two types of cues (positive vs. negative) at two alternated times (time 1 vs. time 2) as a function of the type of time 2 presentation (in isolation vs. simultaneous). Original scores range from −3 (negative) to +3 (positive). The column 'Score' represents the difference between time 1 and time 2

	Positive			Negative		
Presentation	**Time 1**	**Time 2**	**Score**	**Time 1**	**Time 2**	**Score**
Face-based impression						
In isolation	1.03 (1.27)	−0.84 (1.29)	1.87 (1.81)	−0.46 (1.27)	1.24 (1.28)	1.70 (1.56)
Simultaneous	1.02 (1.20)	−0.72 (1.26)	1.73 (1.86)	−0.35 (1.26)	1.42 (1.09)	1.77 (1.49)
Voice-based impression						
In isolation	1.29 (1.04)	−0.59 (1.35)	1.88 (1.63)	−0.55 (1.46)	0.80 (1.44)	1.36 (1.74)
Simultaneous	1.37 (1.04)	0.40 (1.40)	0.97 (1.57)	−0.65 (1.42)	0.13 (1.44)	0.78 (1.73)

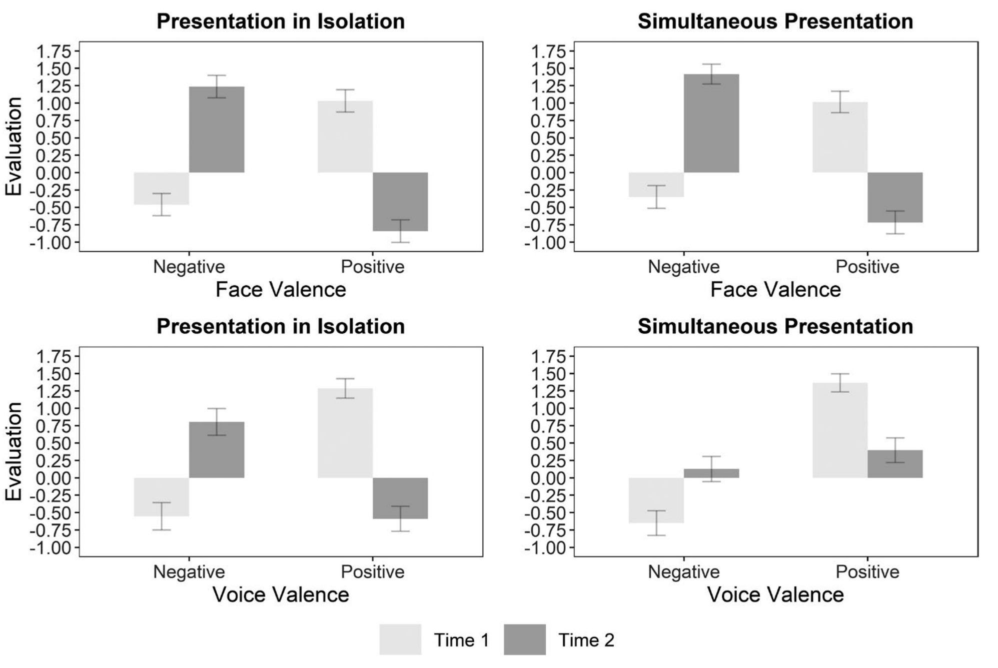

FIGURE 3 Bar graphs of Experiment 3 depicting evaluations given at time 1 (light grey bar) and time 2 (dark grey bar) as a function of first information valence (negative vs. positive) and second information presentation type (in isolation vs. simultaneous). The upper panels refer to face-to-voice impression updating, meaning that first impressions formed on faces were then updated by voices. The lower panels refer to voice-to-face impression updating, meaning that first impressions formed on voices were then updated by faces. Note that the additional information is always of opposite valence (i.e., inconsistent) compared to the first information. Evaluations are expressed on the original scale ranging from −3 (negative) to +3 (positive)

(first information valence: positive vs. negative) × 2 (first information type: face vs. voice) mixed ANOVA on the impression updating scores calculated as in the prior experiments (see Figure 4). The effect of valence of the first information was not significant, $F(1, 246) = 0.13$, $p = .72$, $d = .05$, 95% CI [−0.08, 0.17]. Critically, the effect of information type was significant, $F(1, 246) = 4.10$, $p = .04$, $d = .26$, 95% CI [0.01, 0.51], meaning that voices updated face-based impressions ($M = 1.34$, $SD = 1.80$) more strongly than faces with voice-based impressions ($M = 0.98$, $SD = 1.79$). The interaction between the two factors was significant, $F(1, 246) = 3.78$, $p = .05$, $\eta_p^2 = .02$, 95% CI [0, 0.06]. Negative face-based impressions were updated more by positive voices ($M = 1.45$, $SD = 1.72$) than voice-based impressions by positive faces (M = 0.81, SD = 1.81), t(473) = 2.80, p = .005, d = .26, 95% CI [0.08, 0.44], whereas updating was similar on positive face-based (M = 1.23, SD = 1.87) and voice-based impressions (M = 1.15, SD = 1.76), t(473) = 0.34 p = .74, d = .03, 95% CI [−0.15, 0.21].

The joint analysis showed that when participants encountered the second cue in isolation, their impression of the target updated to greater extent when transitioning from face to voice than vice versa, especially when the new information was positive (and the first cue was negative). However, one should not ignore that the effects tended to slightly diverge across the two experiments considered in isolation. Thus, further empirical replications in support of our findings are encouraged.

GENERAL DISCUSSION

Our impressions of others can be influenced by both their faces and their voices. Prior research has investigated faces and voices as independent sources of impression (McAller & Belin, 2018; Todorov et al., 2015). Only recently, their cross-modal integration gained new attention (Hansen et al., 2017a; Mileva et al., 2018). Yet, little is known about the face–voice dynamics in impression updating, especially from an interpersonal perspective. Across three studies, we investigated the cross-modal updating of face-based and voice-based impressions. In Experiment 1, we demonstrated that impressions based on faces are updated after hearing voices of opposite valence. In Experiments 2 and 3, we compared the impact of voices and faces to update impressions generated by the alternative cue. We found that a target's voice led to stronger updating of face-based impressions than the other way around. When the first and second cues were presented one at a time (i.e., in isolation), the updating of impressions was moderated by valence, with the greatest updating emerging for positive voices over negative face-based impressions. When the first cue co-occurred with the second information (i.e., simultaneous presentation), the superior impact of the voice (vs. the face) in updating face- (vs. voice-) based impressions generalized across valence.

Our findings provide novel insights into cross-modal impression updating. Prior studies suggested that faces and voices could contribute differently to the ultimate impression that we form about other individuals (Hansen et al., 2017a; Mileva et al., 2018). Whereas some of these studies focused on the impact of two cues when presented simultaneously (Rezlescu et al., 2015), less was known about their interplay when one cue follows the alternative one (i.e., impression updating). In the only research considering faces and voices in sequential presentation (Hansen et al., 2018), a rapid (few milliseconds) sequential presentation of the two cues served to make it clear that both belonged to the same target's identity. Such a procedure prevented one from determining the individual contribution of each cue to impression updating. Instead, we opted for a different approach in which faces and voices were presented separately and names were used to link each cue to their respective identity, favouring the measurement of each cue's individual contribution to the two stages of impression formation. This allowed us to show not only that face-based impression can change when people are presented with the voice of the target at a later stage, but also that the updating was stronger when voices (vs. faces) followed the presentation of faces (vs. voices). Indeed, our findings suggest that impressions tend to be more malleable when transitioning from the view (e.g., a picture) of the face to the mere listening of the voice (e.g., on the phone). Whereas the effect in Experiment 2 could be partly attributed to the fact that negative faces led to more negative impressions than negative voices, Experiment 3 ruled out this possibility: even though both positive and negative faces and voices led to comparable first impression, the role of either cue in updating such impressions showed different patterns. Eventually, a joint analysis of the two experiments found voices to produce a stronger updating than faces. We also found voices to be even superior in updating prior impressions when following an alternative cue presentation (i.e., co-occurrence of both cues at the later stage). This extends literature showing that voice matters more when cues of social groups are concerned (Hansen et al., 2018; Rakić et al., 2011). Therefore, impressions coming from the face could be even more malleable when preceding the simultaneous processing of both

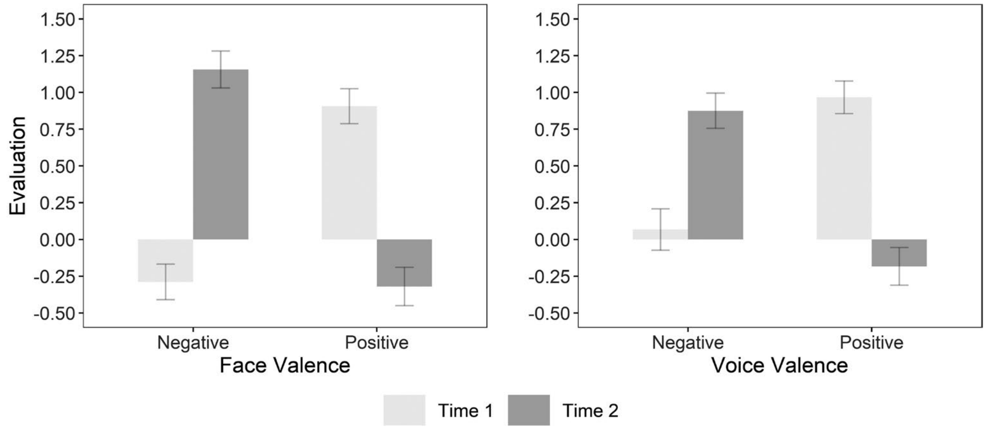

FIGURE 4 Bar graphs of the joint analysis of Experiments 2 and 3 depicting evaluations given at time 1 (light grey bar) and time 2 (dark grey bar) as a function of first information valence (negative vs. positive). The left-hand panel refers to face-to-voice impression updating, meaning that first impressions formed on faces were then updated by voices. The right-hand panel refers to voice-to-face impression updating, meaning that first impressions formed on voices were then updated by faces. Note that the additional information is always of opposite valence (i.e., inconsistent) compared to the first information. Evaluations are expressed on the original scale ranging from −3 (negative) to +3 (positive)

cues (e.g., on a video call). Thus, our findings brought evidence in favour of the superiority of voices for impression updating.

Cross-modal impression updating from faces and voices has so far been mostly investigated within the framework of the intergroup relations (Hansen et al., 2018). The faces and voices' potential to alter impressions formed via the alternative cue was investigated by focusing on one unique feature (i.e., ethnicity) manipulated within each cue and in specific contexts (e.g., recruitment). Studying the dynamics of impressions influenced by non-behavioural cue is crucial also at a more general level and outside specific contexts. Our manipulation of both faces and voices did not confine to one single characteristic of either cue (e.g., skin colour or accent) that could drive inferences on one (intergroup) dimension. Rather, we focused on the overall power of faces and voices, manipulated to be either negative or positive, to form and update global impressions. Hence, our findings offer empirical evidence to the transformative power of both cues on global impressions. Whereas faces and voices were quite comparable on first impressions, the latter showed stronger in updating such impressions. Therefore, our research further showed that interpersonal interactions are modelled by the face–voice interaction contributing to contemporary research about person perception (i.e., Freeman et al., 2020).

Our findings open to a series of new empirical questions. First, while showing the differential impact of faces and voices on updating first impressions, this work did not test the variables that might explain such an effect. For instance, research has shown that the diagnostic value of the second piece of information is key to revise first impressions based on the initial cue (Cone et al., 2017; Skowronski & Carlston, 1989; Rusconi et al., 2020; see also Brambilla et al., 2019, 2021). Diagnosticity can be conceived as the relative weight attributed to a given piece of information within specific environmental conditions. We might speculate that, within our experimental scenarios, voices were perceived as more diagnostic (i.e., more revelatory of a person's true nature) than faces. Future studies could investigate this issue in different ways. First, diagnosticity could be measured at the level of the perceiver: perceivers who place greater weight on faces should behave differently from those who are more sensitive to someone's voice. Alternatively, the diagnosticity of either cue can be experimentally manipulated, for instance by assessing cross-modal impressions in a context where participants are induced to believe that voice-based (vs. face-based) information leads to erroneous judgements.

Moreover, we advise that faces and voices may not share the same salience in person perception (Taylor et al., 1978, 1979). With reference to our stimuli, one could argue that voices, compared to static faces, were more critical for the last evaluation not only because of the quality but also because of the 'quantity' of the information provided. Most studies on voice perception used sub-second recordings of vowels, syllables, or words which may approximate the amount of information extracted from a static face (McAller et al., 2014; O'Connor & Barclay, 2017). We employed instead longer sentences uttered by speakers which could accost the richness of information provided by dynamic faces (Hehman et al., 2015; but see Kościński, 2013; Rhodes et al., 2011). Future studies could address whether changing the nature of the cues (i.e., static vs. dynamic) would alter our findings.

We limited our examination to male targets. Faces and voices exert different effects on social traits' perception depending on the target gender (Fraccaro et al., 2010; Jones et al., 2010). Future research could manipulate the target gender and reveal which non-behavioural cue counts more. Moreover, we employed manipulated faces and voices: despite being common practice in person perception because of large controllability over interstimuli variability, artificial cues could lead to different effects when compared to real cues (Hehman et al., 2017). Future research should address this issue either using unmanipulated faces and voices or controlling for their artificialness. Eventually, our investigation focused on a general evaluation of the targets. Evaluations given on alternate dispositional traits (i.e., trustworthiness, competence, dominance) may bring to differential outcomes in terms of weight of facial and vocal cues (Mileva et al., 2018; Rezlescu et al., 2015). Thus, it could be important to examine whether cross-modal trait-based versus global impression updating yield contrasting results.

In conclusion, our research supports the idea of dynamical cross-modal impression development. We showed that impressions can be updated in a cross-modal fashion, and voices play a more significant role in influencing early evaluations. This is a further step towards understanding how individuals form and update cross-modal impressions, which are at the basis of social interactions. Understanding whether and how cross-modal cues interact and drive impressions allows us to understand better phenomena such as avoidance, mating, and dating. Nowadays, the very first contact people have is often through social networks where visual information is promptly available: others' voice usually comes at a later stage but, according to our findings, it has the 'power' of changing social dynamics. Concluding, these findings have the virtue of shedding light on an additional feature of our talent of navigating the multi-modal social world.

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Matteo Masi: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Simone Mattavelli:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Fabio Fasoli:** Conceptualization (equal); Investigation (equal); Methodology (equal); Writing – review & editing (equal). **Marco Brambilla:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://osf.io/up62v/.

DATA AVAILABILITY STATEMENT

All the materials, data, and analysis code are available at https://osf.io/up62v/.

ORCID

Matteo Masi https://orcid.org/0000-0001-6436-1663
Simone Mattavelli https://orcid.org/0000-0002-8934-8016
Fabio Fasoli https://orcid.org/0000-0002-8043-1630
Marco Brambilla https://orcid.org/0000-0002-4774-3309

REFERENCES

Ambady, N., Bernieri, F., & Richeson, J. (2000). Toward a histology of social behavior: Judgemental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, *32*, 201-271. https://doi.org/10.1016/S0065-2601(00)80006-4

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology*, *99*(2), 207–220. https://doi.org/10.1080/00224545.1976.9924774

Audacity Team. (2019). *Audacity(R): Free Audio Editor and Recorder* (Version 2.3.2) [Computer software]. Retrieved from https://www.audacityteam.org/download/

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64–73. https://doi.org/10.1016/j.jesp.2019.01.003

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, *64*, 187–262. https://doi.org/10.1016/bs.aesp.2021.03.001

Brooks, J. A., & Freeman, J. B. (2018). Psychology and neuroscience of person perception. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., Vol. *4*). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn413

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–543. https://doi.org/10.1016/j.tics.2007.10.001

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265.

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

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467. https://doi.org/10.1016/j.tics.2003.08.014

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology*, *8*(3), 447–461. https://doi.org/10.1177/147470491000800311

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

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Hansen, K., Rakić, T., & Steffens, M. C. (2017a). Competent and warm? *Experimental Psychology*, *64*(1), 27–36. https://doi.org/10.1027/1618-3169/a000348

Hansen, K., Rakić, T., & Steffens, M. C. (2018). Foreign-looking native-accented people: More competent when first seen rather than heard. *Social Psychological and Personality Science*, *9*(8), 1001–1009. https://doi.org/10.1177/1948550617732389

Hansen, K., Steffens, M. C., Rakić, T., & Wiese, H. (2017b). When appearance does not match accent: Neural correlates of ethnicity-related expectancy violations. *Social Cognitive and Affective Neuroscience*, *12*(3), 507–515. https://doi.org/10.1093/scan/nsw148

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, *41*(8), 1123–1134. https://doi.org/10.1177/0146167215591495

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529. https://doi.org/10.1037/pspa0000090

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*(1), 1–55. https://doi.org/10.1016/0010-0285(92)90002-J

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57–62. https://doi.org/10.1016/j.anbehav.2009.10.003

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: A review. *International Journal of Speech Technology*, *15*(2), 99–117. https://doi.org/10.1007/s10772-011-9125-1

Kościński, K. (2013). Perception of facial attractiveness from static and dynamic stimuli. *Perception*, *42*(2), 163–175. https://doi.org/10.1068/p7378

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), 1–9. 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

McConnell, A. R., Rydell, R. J., Strain, L. M., & Mackie, D. M. (2008). Forming implicit and explicit attitudes toward individuals: Social group association cues. *Journal of Personality and Social Psychology*, *94*(5), 792–807. https://doi.org/10.1037/0022-3514.94.5.792

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 128–138. https://doi.org/10.1037/xhp0000439

O'Connor, J. J. M., & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, *38*(4), 506–512. https://doi.org/10.1016/j.evolhumbehav.2017.03.001

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Pisanski, K., Jones, B. C., Fink, B., O'Connor, J. J., DeBruine, L. M., Röder, S., & Feinberg, D. R. (2016). Voice parameters predict sex-specific body morphology in men and women. *Animal Behaviour*, *112*, 13–22. https://doi.org/10.1016/j.anbehav.2015.11.008

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology*, *100*(1), 16–29. https://doi.org/10.1037/a0021522

Rakić, T., Steffens, M. C., & Sazegar, A. (2020). Do people remember what is prototypical? The role of accent–religion intersectionality for individual and category memory. *Journal of Language and Social Psychology*, *39*(4), 476–494. https://doi.org/10.1177/0261927X20933330

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

Rhodes, G., Lie, H. C., Thevaraja, N., Taylor, L., Iredell, N., Curran, C., Tan, S. Q. C., Carnemolla, P., & Simmons, L. W. (2011). Facial attractiveness ratings from video-clips and static images tell the same story. *PLoS One*, *6*(11), 1–6. https://doi.org/10.1371/journal.pone.0026653

Riva, P., Brambilla, M., & Vaes, J. (2016). Bad guys suffer less (social pain): Moral status influences judgments of others' social suffering. *British Journal of Social Psychology*, *55*, 88-108. https://doi.org/10.1111/bjso.12114

Rule, N. O., & Sutherland, S. L. (2017). Social categorization from faces: Evidence from obvious and ambiguous groups. *Current Directions in Psychological Science*, *26*(3), 231–236. https://doi.org/10.1177/0963721417697970

Rule, N. O., Tskhay, K. O., Brambilla, M., Riva, P., Andrzejewski, S. A., & Krendl, A. C. (2015). The relationship between anti-gay prejudice and the categorization of sexual orientation. *Personality and Individual Differences*, *77*, 74–80. https://doi.org/10.1016/j.paid.2014.12.038

Rusconi, P., Sacchi, S., Brambilla, M., Capellini, R., & Cherubini, P. (2020). Being honest and acting consistently: boundary conditions of the negativity effect in the attribution of morality. *Social Cognition*, *38*(2), 146–178. https://doi.org/10.1521/soco.2020.38.2.146

Scherer, K. R. (1979). Voice and speech correlates of perceived social influence in simulated juries. In K. R. Scherer, H. Giles, & R. S. Clair (Eds.), *The social psychology of language* (pp. 88–120). Blackwell.

Shen, X., & Ferguson, M. J. (2021). How resistant are implicit impressions of facial trustworthiness? When new evidence leads to durable updating. *Journal of Experimental Social Psychology*, *97*, 104219. https://doi.org/10.1016/j.jesp.2021.104219

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*, 103888. https://doi.org/10.1016/j.jesp.2019.103888

Shore, L. M., Chung-Herrera, B. G., Dean, M. A., Ehrhart, K. H., Jung, D. I., Randel, A. E., & Singh, G. (2009). Diversity in organizations: Where are we now and where are we going? *Human Resource Management Review*, *19*(2), 117–133. https://doi.org/10.1016/j.hrmr.2008.10.004

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142. https://doi.org/10.1037/0033-2909.105.1.131

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages. *PLoS One*, *10*(7), e0128882. https://doi.org/10.1371/journal.pone.0128882

Taylor, S. E., Crocker, J., Fiske, S. T., Sprinzen, M., & Winkler, D. (1979). The generalizability of salience effects. *Journal of Personality and Social Psychology*, *37*(3), 357–368. https://doi.org/10.1037//0022-3514.37.3.357

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, *11*, 249–288. https://doi.org/10.1016/S0065-2601(08)60009-X

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210–216. https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724–738. https://doi.org/10.1037/a0032335

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

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*(3), 237–242. https://doi.org/10.1177/0963721416683996

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zuckerman, M., & Sinicropi, V. (2011). When physical and vocal attractiveness differ: Effects on favorability of interpersonal impressions. *Journal of Nonverbal Behavior*, *35*(2), 75–86. https://doi.org/10.1007/s10919-011-0106-5

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Masi, M., Mattavelli, S., Fasoli, F., & Brambilla, M. (2021). Cross-modal impression updating: Dynamic impression updating from face to voice and the other way around. *British Journal of Social Psychology*, 00, 1–18. https://doi.org/10.1111/bjso.12511