Consciousness and Cognition 88 (2021) 103072

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Vocal signals only impact speakers' own emotions when they are self-attributed

Louise Goupil[a,b,*], Petter Johansson[c], Lars Hall[c], Jean-Julien Aucouturier[a]

[a] *STMS UMR 9912 (CNRS/IRCAM/SU), Paris, France*
[b] *University of East London, London, UK*
[c] *Lund University Cognitive Science, Lund University, Lund, Sweden*

ARTICLE INFO

Keywords:
Emotions
Feelings
Vocal signals
Self-perception

ABSTRACT

Emotions are often accompanied by vocalizations whose acoustic features provide information about the physiological state of the speaker. Here, we ask if perceiving these affective signals in one's own voice has an impact on one's own emotional state, and if it is necessary to identify these signals as self-originated for the emotional effect to occur. Participants had to deliberate out loud about how they would feel in various familiar emotional scenarios, while we covertly manipulated their voices in order to make them sound happy or sad. Perceiving the artificial affective signals in their own voice altered participants' judgements about how they would feel in these situations. Crucially, this effect disappeared when participants detected the vocal manipulation, either explicitly or implicitly. The original valence of the scenarios also modulated the vocal feedback effect. These results highlight the role of the exteroception of self-attributed affective signals in the emergence of emotional feelings.

1. Introduction

Emotions are multidimensional states, with both physiological (e.g., increased breathing and heart rate), expressive (e.g., smiles, frowns, growls) and subjective components (e.g., happiness, sadness, fear). While this conceptualization of emotions is relatively uncontroversial, the order in which each of these components occur within an emotional episode still remains a matter of debate. Contrary to the common intuition that bodily changes follow subjective feelings, psychologists have long hypothesized that in fact, feelings result from the perception of physiological changes that arise in the body in response to an event (James, 1890; Laird & Lacasse, 2014; Schachter & Singer, 1962). As William James famously stated, although "*common-sense says, we loose our fortune, are sorry and weep*, […] *the more rational statement is that we feel sorry because we cry*" (James, 1890). While they differ in their formalization of how afferent bodily signals are integrated by the brain to give rise to conscious emotional experiences, most modern theories of emotions still highlight this as an important aspect of emotional experience (Barrett, 2017; Damasio, 2003; LeDoux & Brown, 2017; Panksepp, 2007). Yet, although a lot of research has been dedicated to this issue, it has proven difficult to empirically confirm whether or not the perception of bodily changes plays a causal role in the emergence of emotional subjective experience.

The main strategy has been to causally manipulate bodily signals (Critchley & Garfinkel, 2017; Laird & Lacasse, 2014; Schachter & Singer, 1962; Strack, Martin, & Stepper, 1988; Wagenmakers et al., 2016). The brain receives bodily signals through two main channels: either exteroceptively (tactile, visual or auditory inputs) or interoceptively (cardiac, intestinal or respiratory inputs), and

[*] Corresponding author at: University of East London, School of Psychology, London, UK.
E-mail address: lougoupil@gmail.com (L. Goupil).

https://doi.org/10.1016/j.concog.2020.103072
Received 28 October 2019; Received in revised form 2 November 2020; Accepted 12 December 2020
Available online 3 January 2021

recent evidence suggests that these two types of signals are processed by a shared network supporting bodily self-consciousness (Park & Blanke, 2019). Both exteroceptive and interoceptive bodily signals are thought to play a role in emotions (Damasio & Carvalho, 2013; Laird & Lacasse, 2014), but their contribution to the emergence of feelings has essentially been investigated in separate lines of research.

A first line of research focused on exteroception, and essentially on the perception of bodily feedback related to facial expressions. Influential studies focusing on proprioception initially provided evidence that modifying agents' facial configuration (e.g., artificially inducing a smile) changes their emotional experience (Laird & Lacasse, 2014; Strack et al., 1988). However, this approach has proven to be quite unreliable, with a recent paper reporting a failure to replicate this effect in a large number of independent laboratories (Wagenmakers et al., 2016). Arguably, the evanescence of the facial feedback effect might be due to the fact that participants are often aware of the manipulation. For instance, recent research showed that being observed, or not, has an impact on the effect (Noah, Schul, & Mayo, 2018). This type of finding is consistent with the idea that self-attribution (i.e., identifying signals as genuinely originating from the self) is a necessary condition for bodily signals to impact emotional experience, an issue to which we come back below.

A second line of research focused on interoception, showing for instance that the cardiac cycle impacts emotional judgements about facial expressions of fear (Critchley & Garfinkel, 2017), and that receiving false feedback of increased heart rate enhances the perceived emotional intensity of neutral faces (Gray, Harrison, Wiens, & Critchley, 2007). By focusing on the perception of external stimuli, rather than on emotional experience per se, this line of research has established a causal impact of interoceptive feedback on the evaluation of social stimuli. However, it remains unclear whether this is accompanied by substantial changes in subjective emotional experience.

Perhaps more importantly, when using manipulation to study the feedback effects of interoceptive and proprioceptive signals, it is difficult to conceal the true nature of the experiment. For example, if you ask participants to hold a pencil in their mouths while rating cartoons, the artificiality of the forced smile may lead at least some participants to determine the purpose of the prop. This is especially critical if, as suggested above, self-attribution is an important prerequisite for bodily signals to genuinely impact emotional feelings.

Consistent with this idea, a third line of research has focused on physiological arousal, and relied on injections of adrenaline or placebos to probe the impact of self-attributions of physiological changes on the intensity of emotional experience (Laird & Lacasse, 2014; Schachter & Singer, 1962). For instance, a classic study suggested that in the absence of information about the effects of adrenalin injections, participants are more likely to incorporate these physiological effects and show strong emotional reactions, compared to when they are informed about the effects of adrenalin (Schachter & Singer, 1962). This suggests that perceiving physiological signs of arousal leads to different emotional effects depending on whether they are perceived as originating from the self, or from an external cause. Yet, it remains unclear at what level of consciousness such distinction occurs. Do bodily changes have to be consciously perceived as externally-originated to break the effect, or is it sufficient that they are implicitly processed as external, even if participants have no conscious awareness of being manipulated?

Here, we ask whether the self-attribution of affective bodily signals is a necessary condition for them to causally impact subjective emotional experience. To overcome the limitations listed above, we test this hypothesis by focusing on the voice, thereby returning to James' initial formulation that "*we feel sorry because we cry*" and not the other way around. Like proprioceptive and interoceptive signals, the voice is deeply modified during emotional episodes, and reflects physiological changes that happen within the body. For instance, happy voices tend to have more energy in high frequencies due to the labial spreading associated with smiling (Ponsot, Arias, & Aucouturier, 2018). Similarly, during states of high arousal, increased muscle tension can cause an increase in subglottal air pressure leading to the production of rough vocalizations (Fitch, Neubauer, & Herzel, 2002). Such affective signals are often assumed to primarily serve a communicative function (Dezecache, Mercier, & Scott-Phillips, 2013; Fitch et al., 2002; Scherer, 2003), but here we hypothesize that these displays, like other types of bodily signals which can be perceived via proprioception and interoception, also play a role in the emergence of the signaler's own emotional experience.

It was recently discovered that vocal signals can be captured and progressively altered without speakers necessarily noticing the manipulation (Aucouturier et al., 2016). This leads to a situation in which artificial affective signals (e.g., an up-ward shift and boost in high frequencies mimicking a happy voice) can be introduced in the voice of a speaker and, at least in some circumstances, perceived as self-originated. In this original study, participants' self-reports revealed that being exposed to artificial affective signals in their own voice altered their mood in directions that were congruent with the manipulation (i.e., perceiving happy vocal signals induced more positive reports in the speaker). However, because it featured a very low rate of manipulation detection (14%), this study did not allow to conclude whether the emotional effect is limited to cases where the manipulated vocal signals are perceived as self-generated.

The ability to distinguish self-produced actions from external events (i.e., self-monitoring) is thought to rely on internal forward models, that allow comparing the perceptual consequences of planned actions against the perceptual feedback stemming from these planned actions (Wolpert & Ghahramani, 2000). Consistent with these models, empirical research has shown that hearing self-produced speech leads to attenuated responses in the auditory cortex compared to externally generated speech (Chang, Niziolek, Knight, Nagarajan, & Houde, 2013). In addition, research relying on speech perturbation has shown that when discrepancies arise (e.g., when the fundamental frequency of the vocal feedback is altered), a neural error signal is generated (Chang et al., 2013; Tian & Poeppel, 2015) which, behaviorally, may lead to an acoustic compensation of the change (Hafke, 2008; Jones & Munhall, 2000). This compensation is thought to reflect the detection of an error in the speech production process (Chang et al., 2013; Jones & Munhall, 2000; Max, Wauacet, & Vincene, 2003).

Following these models, we could therefore take the presence of acoustic compensation during emotional vocal feedback as evidence of implicitly detecting that the vocal signal is not self-generated, which allows us to ask whether such detection leads to a dismissal of the affective signals it contains.

As mentioned above, the original emotional vocal feedback procedure featured both very low rates of explicit detection, and no evidence of acoustic compensation (Aucouturier et al., 2016). Here, we rely on a more demanding procedure, involving frequent reversals of the vocal feedback manipulation in a within-participant paradigm. This new procedure induces a greater rate of explicit and implicit detections of the experimental manipulation, and notably allows us to detect possible acoustic compensation at the level of the participant. Using this new procedure, we asked participants to deliberate out loud about how they would feel in familiar scenarios that systematically varied in valence and arousal (e.g., receiving good or bad news). While the participants spoke, we covertly manipulated subtle acoustic features (fundamental frequency and timbre) in their voices in order to introduce affective signals typically associated with happiness or sadness (Fig. 1). We predicted that covert manipulation of affective signals in their own voice should impact participants' emotional experience in the congruent direction (e.g., happy signals should induce positive feelings). Additionally, we predicted that this effect should disappear when speakers detected our manipulation, either explicitly (as assessed by progressive questions during debriefing) or implicitly (as evidenced through vocal compensation).

2. Materials and methods

2.1. Experimental design

In order to assess the impact of self-attributed vocal affective signals on emotional experience, participants read 36 emotional scenarios out loud, and then deliberated out loud about how they think they would feel in this situation. After reading and describing how they would feel (which took on average 79 s ± 42 SD), participants had to summarize their assumed emotional state using: 1) a continuous valence scale ranging from "very negative" (0) to "very positive" (100); 2) a continuous arousal scale ranging from "not aroused" (0) at all to "very aroused" (100), and 3) to report how sure they were that they would feel this way on a continuous confidence scale ranging from "not confident at all" (0) to "very confident" (100). While they spoke, participants' voices were covertly manipulated to make them sound *happy* (i.e., higher fundamental frequency, and boost in high-medium frequencies mimicking the acoustic consequences of smiling, Ponsot, Arias, & Aucouturier, 2018) in 1/3 of the trials, *sad* (i.e., lower fundamental frequency, darker spectrum) in 1/3 of the trials, or remained unchanged (i.e., *neutral* condition) in the remaining 1/3 of the trials. The covert manipulation started as soon as participants began reading the scenario out loud, and increased progressively with a ramp of 2.5 s to reach pre-defined values for each acoustic manipulation (fundamental frequency and vocal timbre). The manipulation stayed on while the participants discussed out loud how they think they would feel in the scenario.

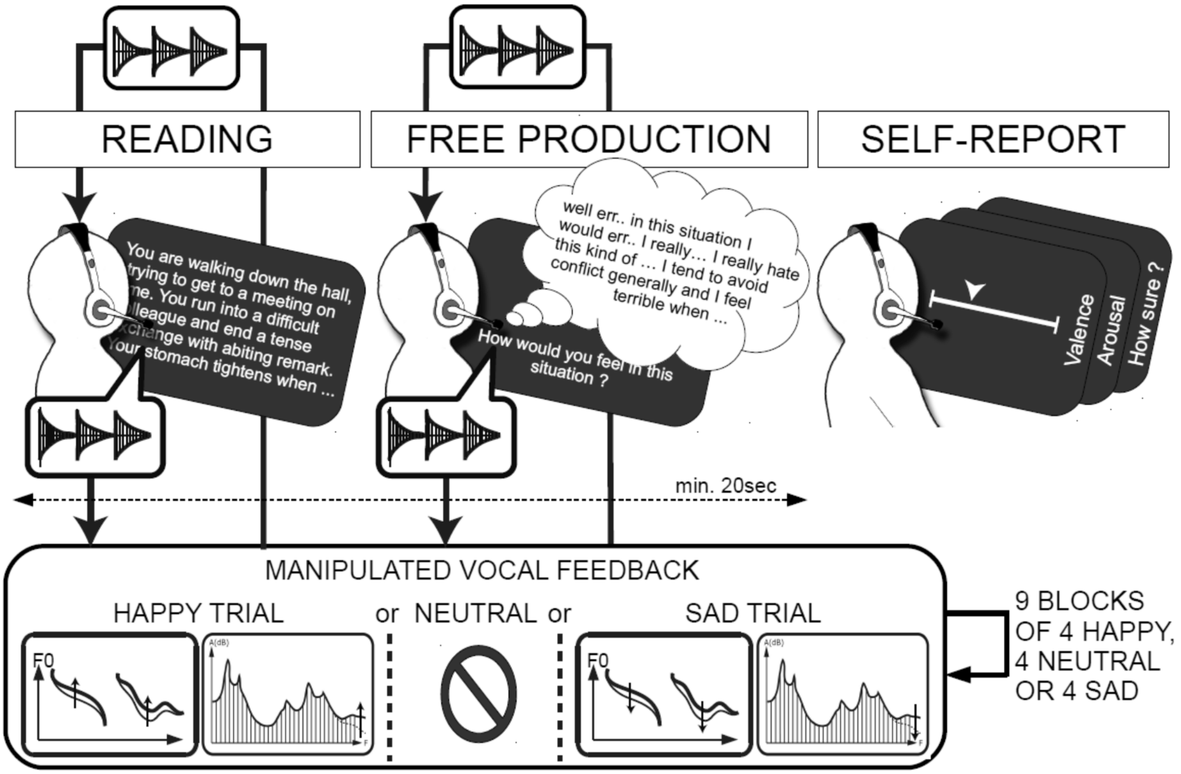
READING
FREE PRODUCTION
SELF-REPORT
You are walking down the hall, rying to get to a meeting on me. You run into a difficult lleague and end a tense xchange with abiting remark. Your stomach tightens when ...
well err.. in this situation I would err.. I really... I really hate this kind of ... I tend to avoid conflict generally and I feel terrible when ...
How would you feel in this situation ?
Valence
Arousal
How sure ?
min. 20sec
MANIPULATED VOCAL FEEDBACK
HAPPY TRIAL
or NEUTRAL or
SAD TRIAL
F0
A(dB)
F
F0
A(dB)
F
9 BLOCKS OF 4 HAPPY, 4 NEUTRAL OR 4 SAD

Fig. 1. Experimental paradigm. Participants read vignettes depicting familiar emotional scenarios, and then deliberated out loud about how they would feel in these situations. They then summarized their emotional state on a valence scale, an arousal scale and a confidence scale. While they spoke, their voice was covertly manipulated to make it sound *happy* (i.e., higher fundamental frequency, brighter spectrum) or *sad* (i.e., lower fundamental frequency, darker spectrum), or remained unchanged (i.e., *neutral*).

Participants were fitted with closed headsets minimizing the contamination from environmental noise and their non-manipulated voice (Beyerdynamics – DT770), and their voice was recorded with a DPA 4088 headset and an RME UCX Fireface sound card, that allows for a roundtrip latency under 20 ms. An Apple MacBook Pro running PsychoPy (Peirce, 2007) was used to control stimulus presentation, recording responses, as well as communicate in real time with the voice transformation software (DAVID) to apply the various voice transformations depending on experimental conditions. Participants were asked to read each of the scenarios out loud, before speaking freely to describe how they would feel in such a situation. To conceal the true nature of the experiment, participants were told that the experiment aimed at uncovering why and how emotional feelings arise in various imaginary situations.

In a previous study using a between-subject design (i.e., each participant heard her/his voice manipulated in only one direction), and a very slow introduction of the vocal-feedback manipulation, we found both that most participants did not detect that their voice had been manipulated, and that none of the participants displayed acoustic compensation (Aucouturier et al., 2016). In the present experiment, we used a design that we deemed likely to result in a higher detection rate, either explicitly or as indexed through acoustic compensation. This time, we used a within-subject design, in which the direction of the vocal-feedback manipulation was changed after every 4 trials, resulting in 9 blocks of 4 *happy*, *sad* or *neutral* trials. Thus, there were 3 non-consecutive blocks of 4 trials for each of the vocal feedback manipulation (hereafter VFM) conditions (12 trials with the *happy* VFM, 12 with the *sad* VFM, and 12 *neutral* trials). Pilot testing revealed that in this within-subject design, the percentage of participants who detected the manipulation increased as opposed to a between-subject design where only one effect was slowly applied to each participants' voice (Aucouturier et al., 2016). Vocal feedback manipulation, block order and scenarios were pseudo-randomized across participants, such that across groups of 18 participants, each scenario appeared in the *happy*, *sad* or *neutral* condition for 1/3 of the participants.

Following the experiment, participants were debriefed with progressive questions, in order to assess whether they detected the vocal feedback manipulation or not. Depending on their responses, they were given a score of detection ranging from 1 to 6 by the experimenter (1 = "*you intentionally manipulated my voice*"; 2 = "*my voice was higher / lower sometimes, and it changed during the experiment*"; 3 = "*my voice sounded strange sometimes, and it changed during the experiment*"; 4 = "*my voice sounded strange, and it was not only because I am not used to hearing myself through headphones*"; 5 = "*my voice sounded strange, but it is probably because I am not used to hearing myself through headphones*" ; 6 = "*my voice was absolutely fine*"). At this stage, the experimenter remained blind to the participant's behavior during the experiment (i.e., did not hear the participant's responses during the experiment, did not know whether the participant compensated acoustically or not, and did not know whether they displayed the emotional effect or not). Participants also filled in a questionnaire in order to assess their level of alexithymia (TAS scale, Zimmermann, Quartier, Bernard, Salamin, & Maggiori, 2007) as well as social anhedonia (SHAPS scale, Gaillard, Gourion, & Llorca, 2013). At the end of the experiment, they were debriefed and informed of the true purpose of the study.

2.2. Participants

Sample size was chosen following pilot testing, which suggested that approximately one third of the participants detected the manipulation explicitly, one third implicitly, and one third showed no detection at all. Given our design which required forming groups of 18 participants (see experimental design above), we thus aimed to test fifty-four participants. In total, fifty-five participants were tested, but seven participants had to be excluded due to technical problems with recording, or real-time vocal transformations (i.e., failure of the software, or communication between the software and the experimental interface coded in python), leaving N = 48 participants in the final sample (29 females; age = 23 ± 3.24 SD).

2.3. Stimuli

36 emotional scenarios were adapted and translated into French from (Wilson-Mendenhall, Barrett, & Barsalou, 2013). These scenarios were originally designed to describe familiar situations able to induce varying experience of valence and arousal in participants. We selected 9 scenarios per valence (positive/negative) and arousal (high/low arousal) quadrants, based on the normative ratings provided by Wilson-Mendenhall et al., (2013) (see supplementary material for four examples of scenarios). We verified in an independent group of 10 participants with equivalent demographic background as our main sample (4 females, age = 28.5 ± 4.48 SD) that the stimuli were perceived similarly in our population (mean valence ratings for positive scenarios on a scale from 0 to 100: M = 69.7 ± 9 SD, negative scenarios: M = 40.2 ± 9.3 SD; arousal ratings for high arousal scenarios on a scale from 0 to 100: M = 60.3 ± 19 SD; low arousal scenarios: M = 48.5 ± 11.1 SD).

2.4. Voice transformation

Subtle affective signals were artificially introduced in participants' speech in real time using the voice transformation technique introduced in Rachman et al. (2018). The software, called DAVID, uses a selection of digital audio effects such as fundamental frequency shifting and spectral filtering to simulate emotional expression in running speech, with a very-low latency compatible with real-time vocal production. For the *happy* Vocal Feedback Manipulation (VFM), we applied a positive fundamental frequency shift (+50 cents) as well as a spectral modification aiming to simulate the impact of smiling (notch filter at 2880 Hz, gain = 3, Q = 0.74) (Ponsot et al., 2018). For the *sad* VFM, we applied a negative fundamental frequency shift (−70 cents) as well as a spectral modification aiming to attenuate the power in high frequencies, resulting in a darker sound (high-shelf filter, 8000 Hz, gain = 0.25, Q = 1) (Aucouturier et al., 2016; Scherer, 2003). For the *neutral* condition, the voice was routed through the same algorithm so that the latency (20 ms) and processing of the voice was the same as in the VFM conditions, except that no transformation was applied.

2.5. Pre-processing and statistical analysis

For each trial and each participant, we used the Praat software (Boersma, 2001) to compute the fundamental frequency from two different recordings of the participant's voice: the non-modified recording (i.e., the natural voice produced by the participant, which is the input of the voice transformation algorithm) and the modified recording (i.e., the artificial voice heard by the participant, which is the output of the voice transformation algorithm). This analysis revealed that, although generally effective, our fundamental frequency transposition algorithm sometimes failed because of the low vocal or audio quality in certain trials (e.g., because the position of the microphone was not adequate). Thus, we excluded trials for which the difference between the fundamental frequency of the modified and non-modified voice did not show the intended transposition (e.g., for the *sad* condition, we excluded trials if the fundamental frequency of the modified voice was not lower than the fundamental frequency of the non-modified voice). This pre-processing lead to excluding 7% of the data.

Trials in the *neutral* VFM condition were then used as a baseline to assess the impact of the *happy* and *sad* VFM on ratings. For each participant, ratings were normalized with respect to the participant's *neutral* trials, following this formula: $zX = (X - m\_neutral)/sd\_neutral$, were zX corresponds to normalized ratings (valence, arousal or confidence), X to data samples, *m_neutral* to the average and *sd_neutral* to the standard deviation of the ratings given by this participant in neutral trials. Over all participants, the average neutral ratings were: valence: 52.95 ± 7.47 SD; arousal: 52.01 ± 8.13 SD; confidence: 79.42 ± 13.05 SD.

Explicit detection was assessed from participants' responses during the debriefing: participants who detected that their voice had been manipulated (i.e., detection scores $<= 4$) constituted the group of explicit detectors, and participants who did not detect the manipulation at all (scores > 4) constituted the group of non-detectors (see Table 1).

Implicit detection was estimated for each participant by assessing whether the fundamental frequency of their non-modified voice deviated from its normal range in each of the two VFM conditions (*happy* and *sad*). We assessed acoustic compensation at the level of the participant (i.e. globally, over all trials of the participant) by testing whether the fundamental frequency extracted from the non-modified recordings in each of the VFM condition (*happy / sad*) differed from the normal range of variation in the participant's fundamental frequency, measured across all conditions. The fundamental frequency of the non-modified recordings of each participant was normalized (z-scored) over all trials of all three conditions of the participant. The z-scored values for the *happy* and *sad* trials of the participant were then tested against zero, with two separate one-sample t-tests, to assess their deviation from the participant's normal range. A significance threshold of <0.1 was adopted in order to avoid missing implicit detection through lack of power at the individual level. Similar results were obtained when using a threshold of $p < 0.05$ to classify individual participants, and also when normalizing with respect to neutral trials only.

This analysis revealed that 50% (N = 24; 16 explicit non-detectors) of the participants did not show any acoustic compensation; 16.7% (N = 8; 5 non-detectors) compensated in the *happy* condition (i.e., produced a lower fundamental frequency when their voice had been transposed upwards); 22.9% (N = 11; 7 non-detectors) compensated in the *sad* condition (i.e., produced a higher fundamental frequency when their voice had been transposed downward); 2% (N = 1 non-detector) compensated in both directions; unexpectedly, 8.3% (N = 4; 2 non-detectors) of the participants displayed anti-compensation in the happy condition (i.e. their voice had a higher than average fundamental frequency in the happy trials).

For each individual participant and condition, we classified all trials of a condition as "compensated" if the participant compensated in that condition (conservatively, the anti-compensations observed in the *happy* condition were also counted as compensation). This resulted in 188 trials (96 for *happy*, 92 for *sad*) in the compensated condition, and 506 trials (276 for *happy*, 230 for *sad*) in the non-compensated condition. Fig. S1 shows an analysis of the fundamental frequency of the non-modified voices for compensated versus non-compensated trials, which confirmed that this classification efficiently separated trials with substantial acoustic compensation from trials without acoustic compensation at the level of the group.

Because of the unbalanced numbers of detectors vs. non-detectors, and compensated vs. non-compensated trials, we used hierarchical linear mixed regressions to assess the statistical significance of our results. Hierarchical linear models were fit with the maximum likelihood method. Normalized valence ratings, confidence ratings, arousal ratings, the non-modified (pre-VFM) and the modified (post-FVM) fundamental frequency served as dependent variables in separate models. Participant and sentence number were entered as random factors. Fixed factors were the VFM condition (*happy* vs. *sad* vs. *neutral*), the explicit detection condition (*detector* vs. *non-detector* as between-participant factor, determined as described above), the detection level (using each of the six levels of the detection scale), or acoustic compensation (*compensated* vs. *not-compensated* trials, as described above). We report beta estimates, standard errors, t-values (with Satterthwaite approximations to degrees of freedom), p-values and chi-squares for hierarchical nested model comparisons with likelihood ratio tests (Gelman & Hill, 2007) with the *lme4* and *lmerTest* packages in *R* (Kuznetsova et al.,

Table 1
Explicit Detection results.

Detection level	Description	Number of participants
1	*you intentionally manipulated my voice*	2
2	*my voice was higher / lower sometimes, and it changed during the experiment*	14
3	*my voice sounded strange sometimes, and it changed during the experiment*	0
4	*my voice sounded strange, and it was not only because I am not used to hearing myself through headphones*	1
5	*my voice sounded strange, but it is probably because I am not used to hearing myself through headphones*	8
6	*my voice was absolutely fine*	23

2014).

3. Results

3.1. Impact of participants' felt emotions on the fundamental frequency of their voice

Before assessing the impact of the Vocal Feedback Manipulation (hereafter VFM) on participants' emotions, we first examined how the fundamental frequency of participants' non-modified voices varied with their emotional reports in the non-manipulated trials (i.e., the *neutral* condition). This procedure allowed us to check that we could replicate previous findings concerning the relationship between the fundamental frequency of speakers' voices, and their self-reported valence and arousal (Aucouturier et al., 2016; Scherer, 2003). More importantly, it also allowed us to examine the relationship between fundamental frequency and confidence in emotional judgements. Previous research has documented that epistemic uncertainty (i.e., doubting about one's own state of knowledge) is typically associated with a rising intonation and reduced volume (Goupil & Aucouturier, n.d.; Jiang & Pell, 2017), but the acoustic correlates of emotional uncertainty (i.e., doubting about one's own emotional state) remain unclear so far.

The fundamental frequency of participants' voices in *neutral* trials was averaged separately depending on whether participants reported feeling a negative or positive emotion (median split of valence ratings, see Fig. 2A), high or low arousal (Fig. 2B) and high or low confidence (Fig. 2C). Trials in which participants reported negative emotions were accompanied by lower fundamental frequency than trials in which they reported positive emotions ($t(47) = 2.17$, $p = 0.035$, $d = 0.44$). Similarly, trials with low confidence on the reported emotion had lower fundamental frequency than trials with high confidence ($t(46) = 2.8$, $p = 0.007$, $d = 0.51$). However, arousal only marginally impacted fundamental frequency, with low arousal corresponding to a lower fundamental frequency than high arousal ($t(47) = 1.96$, $p = 0.056$, $d = 0.37$).

3.2. Impact of the vocal feedback manipulations on participants' felt emotions – whole group

Next, we turned to our main research question, and examined whether the VFM had a main effect on the valence, arousal and confidence ratings provided by the participants, and whether this effect was congruent with the relationship between participants' felt emotions and the fundamental frequency of their voice reported above. Over the whole group of participants (detectors and non-detectors, compensators and non-compensators), a repeated measure multi-variate analysis of variance (rm-MANOVA) revealed a significant effect of the VFM condition (*happy / neutral / sad*) on our three main measures of valence, arousal and confidence ratings (F (2,94) = 2.89, Pillai's trace = 0.17, p = 0.01). This first analysis at the level of the whole group of participants, and over the three dimensions of emotional feelings explored in this study, is compatible with our first hypothesis that VFM should impact emotional feelings.

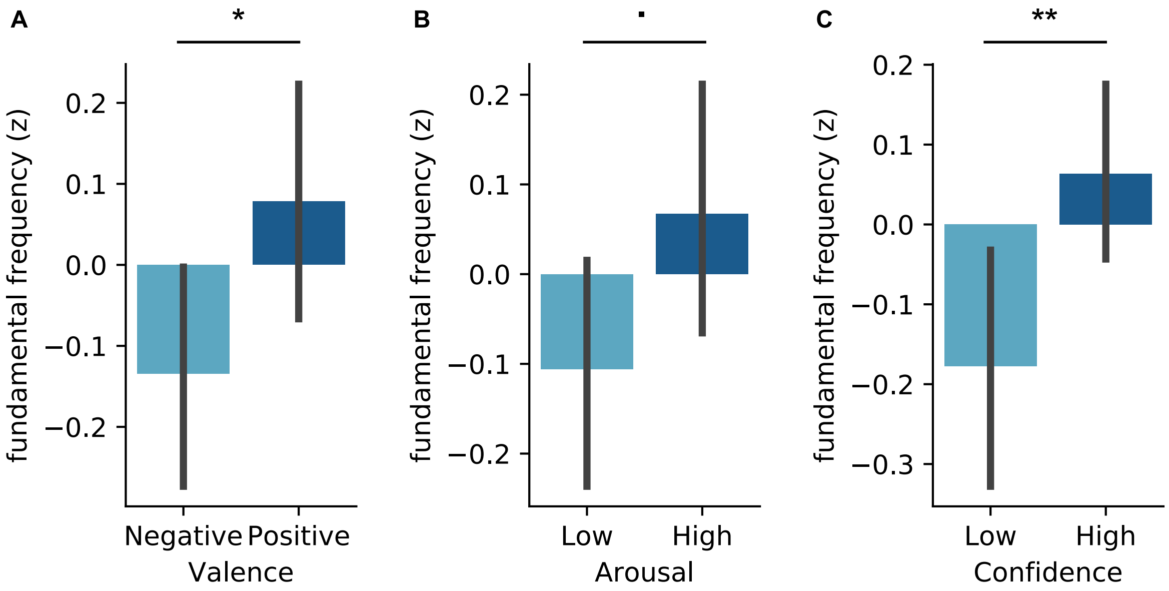

Fig. 2. Impact of felt emotions on the voice. Fundamental frequency of the voices in neutral (i.e., non-modified) trials depending on whether participants reported feeling A) a negative or positive emotion (median split); B) high or low arousal and C) high or low confidence (right panel). Error bars show 95% confidence intervals, and black asterisks the significance of paired t-tests comparisons with: . $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

3.3. Impact of explicit detection on the vocal feedback emotional effect

We assessed explicit detection through a progressive debriefing at the end of the experiment (see methods for details). As can be seen in Table 1, out of the 48 participants, 31 (65%) did not detect the manipulation explicitly (i.e., detection score of 6 *"my voice was absolutely fine"*: N = 23 ; or 5 *"my voice sounded strange, but it is probably because I am not used to hearing myself through headphones"*: N = 8), and 17 (35%) participants detected that their voice had been manipulated (i.e., detection score of 4 *"my voice sounded strange, and it was not only because I am not used to hearing myself through headphones":* N = 1; 2 *"my voice was higher / lower sometimes, and it changed during the experiment"*: N = 14; or 1 *"you intentionally manipulated my voice"*: N = 2). Of note, 23 participants reported that *"their voice was absolutely fine"*, even though they were given the opportunity to respond *"my voice sounded strange, but it is probably because I am not used to hearing myself through headphones"*, an option chosen by a comparatively low number of participants (N = 8). This indicates that roughly half or the participants had no problems with self-identifying their voice in this experiment, despite the fact that it was heard through headphones, which necessarily transforms the vocal feedback they normally hear through the combination of air and bone conduction (Pörschmann, 2000).

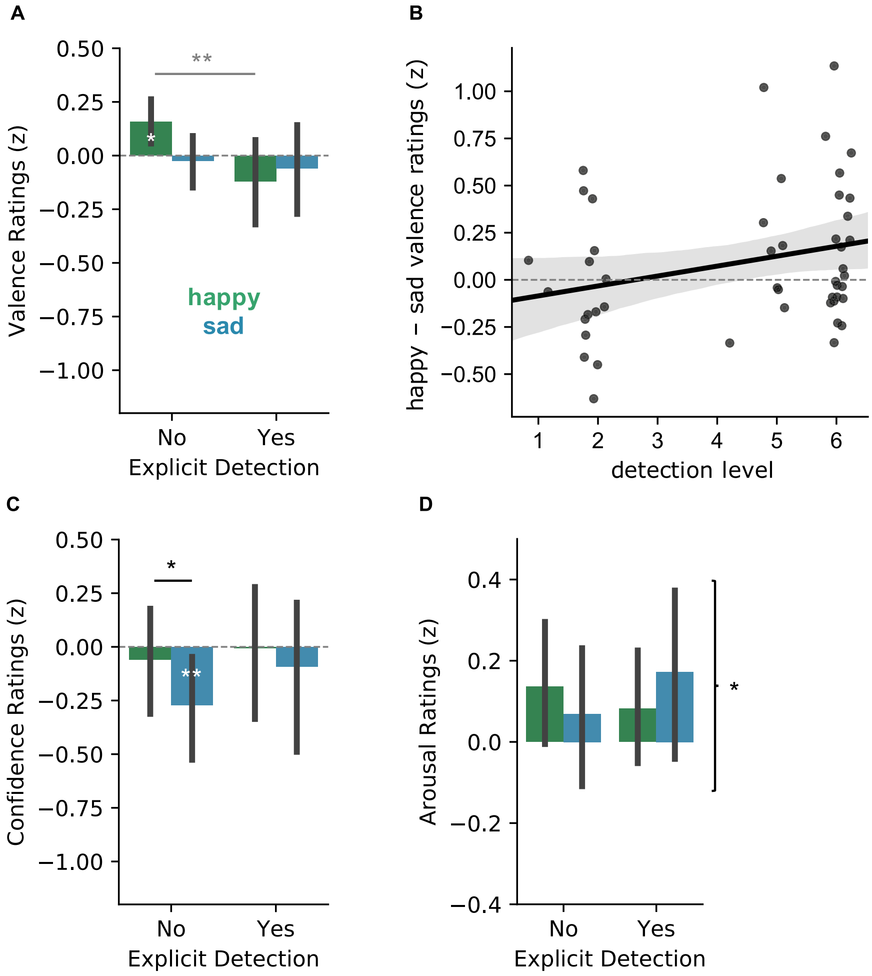

Fig. 3. Main results depending on Explicit Detection. (A) Valence ratings (z-scored) depending on the vocal-feedback manipulation (VFM; *happy*: green, *sad*: blue) and explicit detection. Ratings were normalized with respect to *neutral* trials, so the *neutral* condition is not shown in the Figure (i.e., normalized ratings in the *neutral* condition were equal to zero). (B) Difference between valence ratings in the *happy* minus *sad* condition as a function of the level of detection. Dots show individual data. We show the best fitting regression line with 95% confidence intervals. (C) Confidence ratings (normalized with respect to neutral trials) depending on the VFM and explicit detection. (D) Arousal ratings (normalized with respect to neutral trials) depending on the VFM and explicit detection. Error bars represent 95% confidence intervals. White asterisks represent significance of the model comparisons between experimental conditions (*happy* and *sad* VFM) and the baseline condition (*neutral*), black asterisks represent significance of the model comparisons between the two experimental conditions within groups, and grey asterisks represent significance of the model comparisons between the two experimental conditions between groups. The bracketed black asterisk represents the paired comparison between all manipulated trials (sad and happy) versus the neutral, non-manipulated (i.e., baseline) condition. * represents $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Crucially, and consistent with our second hypothesis, the main effect of the VFM on valence, confidence and arousal reports was driven by the 31 participants that did not explicitly detect the vocal feedback manipulation (see Fig. 3). The interaction between explicit detection (*yes/no*) and the VFM condition (*happy/neutral/sad*) was tested on our three measures separately (valence, confidence and arousal normalized ratings), in hierarchical linear mixed regressions analyses including participant number and sentence number as random factors.

For valence ratings, there was no main effect of the VFM condition (X2 = 1.47, p = 0.48), but an interaction between the VFM and explicit detection (X2 = 7.3, p = 0.026, see Fig. 3A). This interaction reflected the fact that in the group of non-detectors, the *happy* VFM significantly increased valence ratings compared to the baseline (i.e., *neutral)* condition (M = 0.16 ± 0.3 SD; 95% CI [0.04, 0.27]; beta = 0.15 ± 0.06 se, t = 2.5, p = 0.014, d = 0.51), and marginally so compared to the *sad* condition (M = −0.02 ± 0.35 SD; 95% CI [−0.15, 0.1]; beta = 0.11 ± 0.06 se, t = 1.8, p = 0.07, d = 0.56, the difference between the *sad* VFM and the *neutral* condition was not significant, beta = −0.04 ± 0.06 se, t = 0.58, p = 0.56, d = 0.07). In the group of detectors, none of these comparisons were significant (all p-values > 0.14 and $d < 0.3$). In addition, valence ratings in the *happy* condition were higher in the group of non-detectors than in the group of detectors (beta = 0.27, se = 0.08, t = 3.2, p = 0.002, d = 0.83). The direction of the effect is consistent with the fact that speakers displayed higher fundamental frequency when they reported feeling positive rather than negative emotions (see Fig. 2A, for the whole group of participants). As can be seen in Fig. 3B, the interaction between the impact of the VFM on valence ratings and explicit detection was also observable when examining detection as a continuous variable: there was a significant interaction between the VFM and detection level (from 1: highest detection to 6: lowest level of detection; X2 = 6.42, p = 0.04). The difference between the *happy* condition and the *neutral* condition increased with detection level (beta = 0.07, se = 0.03, t = 2.5, p = 0.012), but this effect was not significant for the *sad* condition (beta = 0.026, se = 0.027, t = 0.96, p = 0.33).

Regarding confidence, there was a significant effect of the VFM (X2 = 8.52, p = 0.014, see Fig. 3C) and no significant interaction between the VFM and explicit detection (X2 = 1.07, p = 0.58). The *sad* VFM significantly decreased confidence ratings (M = −0.27 ± 0.73 SD; 95% CI [−0.54, −0.001]) compared to both *neutral* (beta = −0.3 ± 0.1 se, t = −2.8, p = 0.005, d = 0.37) and *happy* VFM conditions (M = −0.06 ± 0.7 SD; 95% CI [−0.32, 0.20], beta = −0.25 ± 0. 11 se, t = −2.28, p = 0.02, d = 0.3) in the group of explicit non-detectors. In the group of explicit detectors, none of the comparisons were significant (all p-values > 0.39 and $d < 0.13$; all other comparisons were non-significant). Again, the direction of the effect was consistent with the fact that participants spoke with a lower fundamental frequency when they reported lower confidence (see Fig. 2C, for the whole group of participants).

Regarding arousal, there was a main effect of the VFM (X2 = 7.9, p = 0.019) and no interaction between the VFM and explicit detection (X2 = 0.25, p = 0.9). As can be seen in Fig. 3D, there was a non-specific effect of the vocal feedback manipulation where both the *sad* (M = 0.11 ± 0.5 SD; 95% CI [0, 0.25]; beta = 0.13 ± 0. 07 se, t = 1.88, p = 0.059, d = 0.22) and the *happy* VFM (M = 0.12 ± 0.4 SD; 95% CI [0, 0.23]; beta = 0.16 ± 0. 07 se, t = 2.2, p = 0.027, d = 0.3) increased arousal ratings compared to *neutral* trial), with no difference between the *happy* and the *sad* conditions (beta = 0.03 ± 0. 07 se, t = 0.4, p = 0.69, d = 0.15). Thus, regardless of the specific manipulation, manipulated trials led to higher arousal ratings. We come back to this aspect in the discussion.

In sum, explicit detection of the vocal feedback manipulation generally removed its impact on emotional self-reports for valence and confidence, which were also the two emotional dimensions that were more tightly associated with changes in the fundamental frequency of participants voices in non-modified trials. This finding suggests that affective vocal signals have to be perceived as self-originated in order to directly influence feelings. To further examine this hypothesis, we then turned to our marker of ‘implicit’ error detection in speech production, by examining acoustic compensation in the subgroup of 31 (explicit) non-detectors.

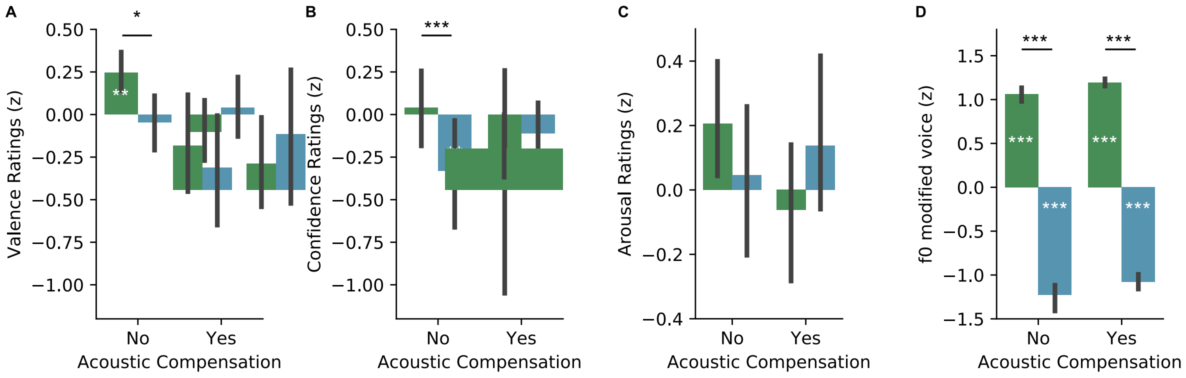

Fig. 4. Main results depending on acoustic compensation (i.e. implicit detection) in the group of explicit non-detectors. (A) Normalized valence ratings, (B) normalized confidence ratings, (C) normalized arousal ratings and (D) fundamental frequency of the modified voice (i.e. output of the VFM) depending on condition (*happy* vs. *sad*) and implicit detection (i.e., acoustic compensation). Ratings were normalized with respect to neutral trials, so the neutral condition (that is therefore equal to zero) is not shown. Error bars represent confidence intervals. White asterisks represent significance of the model comparisons between experimental conditions (*happy* and *sad* VFM) and the baseline condition (*neutral*), black asterisks represent significance of the model comparisons between the two experimental conditions. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.4. Implicit detection of the VFM indexed by acoustic compensation

At the level of the group, we observed substantial acoustic compensation, with a significant main effect of the VFM on the fundamental frequency of the voice produced in manipulated trials (X2 = 44, p < 0.001) and no significant interaction between the VFM and explicit detection (X2 = 4.2, p = 0.12). To document the specific aspect of implicit detection, over and beyond explicit detection, we now focus on non-detectors only (see Fig. S1A for details in the group of explicit detectors). As can be seen in Fig. S1A, the *happy* VFM lead the non-detectors to lower their voice compared to *neutral* trials (M = −0.20 ± 0.36; 95% CI [−0.36, -0.001]; beta = 0.18 ± 0.07 se, t = 2.6, p = 0.009, d = 0.57), while the *sad* VFM lead them to raise their voice (M = 0.27 ± 0.45; 95% CI [0.1, 0.44]; beta = −0.29 ± 0.073 se, t = 3.96, p < 0.001, d = 0.6). At the individual level, we found that out of the 31 participants who failed to explicitly detect the VFM, 15 of them (31% of the total sample of N = 48) showed substantial acoustic compensation for one or both of the VFM conditions (i.e., the fundamental frequency of their voice deviated from their usual range when the *happy* or *sad* VFM was applied, see methods for details). For each participant, trials in a given condition were classified into two “implicit detection” conditions: *compensated* if the participant showed substantial acoustic compensation for that condition over the whole experiment, and *non-compensated* if the participant did not show acoustic compensation for that condition over the whole experiment.

3.5. Impact of implicit detection on the vocal feedback emotional effect

To assess whether implicit detection impacted the effect of the VFM, we ran hierarchical linear mixed regressions on our three measures separately (valence/confidence/arousal normalized ratings) with the VFM condition (*happy/neutral/sad*) and acoustic compensation (*yes/no*) as fixed factors, and participant number and sentence number as random factors.

Regarding valence ratings (Fig. 4A; also see Fig. S2A for histograms allowing to visualize the distribution of the data), there was a marginal effect of the VFM condition (X2 = 5.17, p = 0.075), no main effect of compensation (X2 = 2.47, p = 0.11) and an interaction between compensation and the VFM (X2 = 6, p = 0.047) reflecting the fact that the impact of the VFM on valence ratings was limited to *non-compensated* trials. In *non-compensated* trials, there was a significant increase in valence ratings in the *happy* VFM condition (M = 0.25 ± 0.27 SD) compared to both *neutral* (beta = 0.25 ± 0.094 se, t = 2.6, p = 0.009, d = 0.93) and *sad* trials (M = −0.04 ± 0.37 SD, beta = 0.22 ± 0.089 se, t = 2.53, p = 0.012, d = 0.92). Ratings in the *sad* condition did not significantly differ from the *neutral* condition (beta = −0.02 ± 0.1 se, t = −0.2, p = 0.83, d = 0.12). By contrast, none of these comparisons were significant in *compensated* trials (*happy* M = −0.1 ± 0.26 SD, vs. *neutral*: beta = −0.1 ± 0.12 se, t = −0.8, p = 0.43, d = 0.38; *happy* vs. *sad* M = 0.04 ± 0.26 SD, beta = −0.135 ± 0.15 se, t = −0.93, p = 0.35, d = 0.58; *sad* vs. *neutral*: beta = 0.03 ± 0.13 se, t = −0.27, p = 0.8, d = 0.15).

Similarly, for confidence ratings there was a significant effect of the VFM (X2 = 10.17, p = 0.006), no main effect of acoustic compensation (X2 = 0.15, p = 0.7), and an interaction between compensation and the VFM (X2 = 5.94, p = 0.05), reflecting the fact that compensation also abolished the impact of the VFM on confidence ratings (see Fig. 4B and Fig. S2B for histograms). In *non-compensated* trials, there was a significant decrease in confidence ratings in the *sad* VFM condition (M = −0.33 ± 0.82 SD) compared to both *neutral* (beta = −0.36 ± 0.14 se, t = −2.6, p = 0.009, d = 0.4) and *happy* trials (M = 0.04 ± 0.54 SD, beta = −0.43 ± 0.13 se, t =

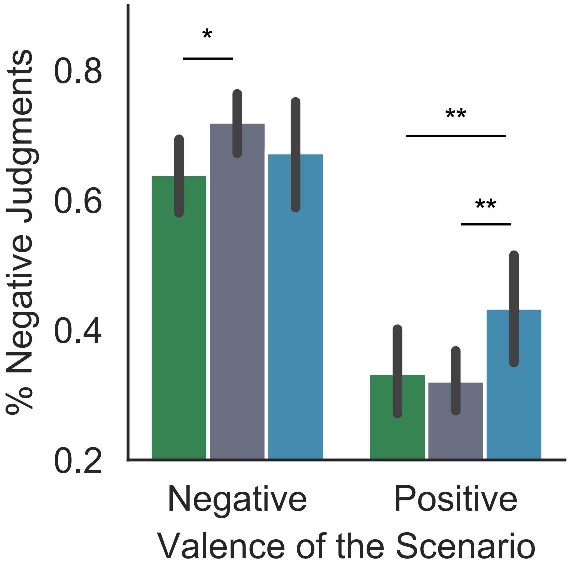

Fig. 5. Interaction between the impact of the VFM on valence ratings and the valence of the context. The percentage of negative responses given by each participant (as defined by a median split of valence ratings) was computed separately in each of the VFM conditions and Valence of the Scenario (green: *happy*, grey: *neutral*; blue: *sad*); * p < 0.05; ** p < 0.01. Comparisons across the positive/negative dimensions are not shown for clarity, but all of them were highly significant (p-values < 0.01). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

−3.4, p < 0.001, d = 0.54). Ratings in the *happy* condition did not significantly differ from the *neutral* condition (beta = 0.07 ± 0.13 se, t = 0.52, p = 0.6, d = 0.07) trials. None of these comparisons were significant in *compensated* trials (*happy* M = −0.34 ± 0.97 vs. *neutral*: beta = −0.34 ± 0.18 se, t = −1.9, p = 0.054, d = 0.35; *happy* vs. *sad,* M = −0.11 ± 0.25: beta = −0.21 ± 0.22 se, t = −0.96, p = 0.33, d = 0.35; *sad* vs. *neutral*: beta = −0.13 ± 0.18 se, t = −0.72, p = 0.5, d = 0.44).

Finally, regarding arousal ratings, there was no interaction between the VFM and acoustic compensation (X2 = 2, p = 0.36, see Fig. 4C), confirming the non-specificity of the impact of the VFM on arousal ratings.

In sum, as was also the case for explicit detection, implicit detection abolished the impact of the VFM on valence and confidence ratings. Importantly, the disappearance of the effect in compensated trials was not trivially due to acoustic compensation somehow counterbalancing the vocal transformations: the fundamental frequency of *modified* voices in *compensated* trials still largely reflected the intended transposition in this subset of the data (see Fig. 4D). In *compensated* trials, the fundamental frequency of the manipulated voices in the *happy* VFM condition (z-scored values, M = 1.19 ± 0.06 SD) remained, as intended, significantly higher than the fundamental frequency of *neutral* voices (M = −0.13 ± 0.19 SD, beta = 1.33 ± 0.05 se, t = 27, p < 0.001) and *sad* manipulated voices (M = −1.08 ± 0.12 SD, beta = 2.3 ± 0.06 se, t = 40, p < 0.001), and the fundamental frequency of the manipulated voices in the *sad* VFM condition remained significantly lower than the fundamental frequency of *neutral* voices (beta = −0.95 ± 0.05 se, t = 19, p < 0.001). This was similar to the pattern observed in non-compensated trials: the fundamental frequency of the manipulated voices in the *happy* VFM condition was significantly higher than the fundamental frequency of *neutral* voices (beta = 1.15 ± 0.04 se, t = 31, p < 0.001), and *sad* manipulated voices (beta = 2.22 ± 0.03 se, t = 64, p < 0.001), and the fundamental frequency of the manipulated voices in the *sad* VFM condition remained significantly lower than the fundamental frequency of *neutral* voices (beta = 1.06 ± 0.04 se, t = 28, p < 0.001). Thus, the impact of the VFM was not abolished in compensators because acoustic compensation canceled our experimental manipulation. Rather, we suggest that acoustic compensation reflected an implicit detection that the affective signals perceived in the voice were not self-generated, leading to a dismissal of these signals during the construction of emotional judgement.

3.6. *Interaction between the vocal feedback effect and the background valence of the scenarios*

Finally, we explored whether the impact of the VFM varied depending on the background valence of the scenarios (Fig. 5). Given that we did not replicate the previous finding that a *sad* VFM should decrease valence ratings, we were particularly interested in whether the valence of the context may impact this effect (Aucouturier et al. 2016 employed a single, neutral text stimulus).

To analyze whether the effect of the VFM depends on the valence of the scenario, we computed for each participant, scenario type (positive/negative) and VFM (*happy*, *neutral*, *sad*) the probability of giving a negative response (determined by a median split on the set of all valence ratings given by each participant). A linear mixed regression including the percentage of negative responses as a dependent variable, the valence of the scenario and the VFM as fixed factors, and participant as a random factor revealed, expectedly, a main effect of the valence of the scenario (X2 = 51.4, p < 0.001), but also an interaction between the VFM and the valence of the scenario (X2 = 9.87, p = 0.0072). Although for positive scenarios the *sad* VFM (M = 0.43 ± 0.2 SD) increased the percentage of negative responses compared to *neutral* trials (M = 0.32 ± 0.12 SD, beta = 0.13 ± 0.05 se, t = 2.75, p = 0.007, d = 0.71) and *happy* trials (M = 0.33 ± 0.17 SD, beta = 0.12 ± 0.04 se, t = 2.84, p = 0.005, d = 0.56), for negative scenarios (M = 0.67 ± 0.23 SD) it did not significantly impact the percentage of negative responses compared to *neutral* trials (M = 0.72 ± 0.12 SD, beta = −0.05 ± 0.04 se, t = −1.1, p = 0.28, d = 0.26), and *happy* trials (M = 0.64 ± 0.17 SD, beta = 0.03 ± 0.037 se, t = −0.89, p = 0.37, d = 0.16). Conversely, the *happy* effect decreased the percentage of negative responses compared to the *neutral* condition for negative scenarios (beta = −0.08 ± 0.04 se, t = −2.1, p = 0.04, d = 0.55), but not significantly so in positive scenarios (beta = 0.009 ± 0.04 se, t = −0.2, p = 0.83, d = 0.08). In sum, the *happy* VFM mostly impacted judgements in negative scenarios, while the sad VFM mostly impacted judgements in positive scenarios.

By contrast, the valence of the scenario did not interact with the effect of the VFM on confidence ratings, in a similar analysis involving the percentage of confident responses (no interaction between the VFM and the valence of the scenario, X2 = 1.1, p > 0.5, see Fig. S3). This speaks against the possibility that the effect of the VFM on confidence ratings was due to perceiving a mismatch between the valence of the scenario and the *sad* vocal feedback effect.

4. Discussion

Covertly introducing artificial affective signals in their voice impacted participants' judgements about how they would feel in various emotional scenarios, but only when they did not detect the experimental manipulation. Such vocal signals are generally assumed to have evolved to serve a communicative function (Dezecache, Mercier, & Scott-Phillips, 2013; Fitch et al., 2002; Scherer, 2003), but here we find that they also play a role for the emergence of conscious emotional states at the individual level.

Applying the happy effect to participants' voices lead them to report more positive emotions. This result is consistent with our initial predictions and with previous research (Aucouturier et al., 2016). By contrast, we did not observe an overall significant decrease in emotional valence in the *sad* condition. This contradicts our prediction that *sad* VFM should lead to a decrease in valence ratings, and no impact on arousal ratings (Aucouturier et al., 2016). Several factors may explain the fact that the *sad* VFM did not significantly impact valence ratings here. In particular, it appears that the impact of the *sad* VFM on emotional judgements was affected by contextual information, and was only manifest in positive scenarios, an issue to which we come back below.

Regarding arousal, we observed a non-specific effect of the VFM, whereby both types of manipulated trials corresponded to higher ratings (Fig. 3). It is possible that, in the context of the present study (reading scenarios out loud), participants did not consistently associate higher or lower fundamental frequency to arousal. This would be consistent with the fact that, contrary to valence, arousal ratings were not strongly associated with the fundamental frequency of participants voices in non-manipulated (neutral) trials (Fig. 2). Instead of an emotional effect, it may have been possible that this increase in arousal ratings reflects physiological arousal linked to the detection of a fundamental frequency mismatch in speech production, regardless of the direction of the change. However, the fact that neither explicit or implicit detection interacted with the increase of arousal makes this possibility unlikely in our view.

The present findings partially replicate a previous study (Aucouturier et al., 2016), but also extend it in several ways, documenting complexities that have important theoretical and methodological implications. First, here we systematically varied the direction of manipulation of the participants' voice within-subject, while the initial study employed a single manipulation which varied between-subjects. In this novel setting, the VFM is sometimes detected by the speakers, either implicitly or explicitly, which allowed us to examine the impact of the detection on the VFM effect. Crucially, we observed that the impact of vocal affective signals depended on these acoustic modifications being perceived as self-generated: although 65% of the participants remained unaware of the manipulation, the substantial number of participants who detected the VFM showed no emotional effect at the group level. We also found that implicit detection that the vocal signal is not self-generated, as defined by acoustic compensation of the manipulation in the participant's non-modified voice, also led to a dismissal of the affective information it contained.

Research has shown that perturbing speech can result in vocal compensation even when participants are not aware of the manipulation (Hafke, 2008; Jones & Munhall, 2000), and that this phenomenon is related to forward-models and self-identification (Max et al., 2003; Tian & Poeppel, 2015). In our initial study on the vocal feedback effect (Aucouturier et al., 2016), we did not observe any acoustic compensation. This is likely due to the fact that we used a continuous stimulation over a long period of time (12 min), and did not change the directionality of the VFM. Here, by contrast, using conditions that most closely resemble the conditions used in speech perturbation studies (Hafke, 2008; Jones & Munhall, 2000; Max et al., 2003; Tian & Poeppel, 2015), we did find substantial acoustic compensation in a subset of our participants. Furthermore, we find that such implicit detection also removes the influence of affective vocal signals on emotional experience. Taken together with past research showing that acoustic compensation reflects the detection of a mismatch between the vocal feedback and internal predictions about the auditory consequences of planned articulation, and that this source monitoring process is linked to perceiving the vocal feedback as originating from the self or not (Jones & Munhall, 2000; Tian & Poeppel, 2015), our finding suggests that only those signals that are identified as originating from the self directly impact felt emotions. This indicates that the identification of bodily signals as originating from the self, potentially through a mechanism of internal forward models, is crucial for them to influence emotional feelings.

One possibility for future studies concerns the factors underlying the individual variability that we observed in detecting the VFM. Here, we did not find any strong association between levels of detection and individual variables such as gender (% of female explicit detectors: 37%; males: 32%), alexithymia (no differences between scores on the TAS scale between detectors and non-detectors, $t(46) = -0.43$, $p = 0.7$) or anhedonia (no differences between scores on the TAS scale between detectors and non-detectors, $t(46) = 0.12$, $p = 0.9$). Further research could further elucidate if there are other individual traits associated with detecting versus assimilating vocal feedback manipulations. We did, however, observe a correlation between alexithymia and the impact of the VFM on emotions (see Fig. S4). This effect is in line with studies reporting differences in the extent to which individuals are able to monitor their own interoceptive and bodily responses (Critchley & Garfinkel, 2017; Laird & Lacasse, 2014), but would have to be confirmed in a bigger sample of participants.

We also observed that the background valence of the scenarios interacted with the VFM emotional effect, with the *happy* VFM mostly impacting negative scenarios, and the sad VFM mostly impacting positive scenarios. Several alternative interpretations can be proposed to account for this observation. First, it is important to note that although we labeled our effects *happy* and *sad* based on previous research on vocal affects (Aucouturier et al., 2016; Scherer, 2003), it is likely that the acoustic signals that we introduced here can actually be interpreted quite differently depending on the context. For instance, neutral states of low arousal such as boredom have also been associated with lower fundamental frequency, while highly negative states of activation such as anger tend to be associated with a higher fundamental frequency (Scherer, 2003). Thus, one possibility is that the "*happy*" and "*sad*" effects actually lead to different inferences depending on the emotion that is afforded by a given scenario. Alternatively, given that we find that the *happy* VFM mostly impacts valence judgements in negative situations, while the *sad* VFM mostly impacts valence judgements in positive situations, it may be that congruent VFM (*happy* on *happy*, *sad* on *sad*) does not warrant an additive effect (for instance, because vocal tone is not informative or surprising in that context), while the same transformations carries more affective information in incongruent situations. The present experiment was not specifically designed to examine these interactions, so follow-up experiments that manipulate vocal signals and situational factors (i.e., whether the scenario depicts a scene that primarily affords joy, weariness, shame, …) orthogonally are required to better understand how the contextual information and vocal signals are integrated to give rise to emotions. Regardless of the precise mechanisms that are involved in these interactions, the present results suggest that, like the interpretation of emotional displays of others (Crivelli & Fridlund, 2018; Wharton, 2009), the interpretation of self-generated signals is subject to an inferential process that integrates contextual cues. This would be consistent with theoretical frameworks proposing that emotional feelings result from integrative inferential processes that integrate bodily signals with other types of information such as episodic and semantic memories (Barrett, 2017; LeDoux & Brown, 2017; Schachter & Singer, 1962).

Another novel aspect of our results is that vocal feedback not only impacted emotional self-reports, but also, how confident participants were that they would feel this way. In neutral trials, we found a lower fundamental frequency in trials where participants provided lower confidence ratings (see Fig. 2C). Congruently, we also found that the *sad* VFM decreased confidence in emotional judgements, thereby suggesting that beyond impacting emotional judgements, the vocal feedback may also play a role in "meta-emotional" judgements (i.e., judgements about emotional judgements). This is consistent with reports showing that confidence in perceptual judgements can be impacted by arousal (Allen et al., 2016), and calls for the development of embodied models of metacognition. It would be especially interesting to test whether a similar vocal feedback effect extends to general epistemic judgements, e.

g. using feedback with modified fundamental frequency to influence a participant's judgements of confidence in non-emotional tasks, such as how certain they are that they see a certain stimulus in a noisy visual search task (see also Goupil, Ponsot, Richardson, Reyes & Aucouturier, 2021).

Our findings have several methodological implications. We have developed a paradigm that allows manipulating participants' vocal feedback in a within subject design, a methodological improvement compared to our previous study (Aucouturier et al., 2016). This will enable further investigations, in particular to study the neural underpinnings of this vocal feedback emotional effect using traditional neuroimaging block paradigms. Our findings also delineate the conditions in which covertly manipulated vocal feedback can impact speakers' emotions: the emotional effect is present only when participants identify the vocal feedback as self-generated. We also observed that the background valence of the context substantially impacts the emotional effect. These things combined have implications for developing vocal feedback applications for clinical practice, like potential non-pharmacological treatments for post-traumatic stress disorder or preoperative anxiety (Guerrier et al., 2019; Scherer et al., 2013).

5. Conclusions

To conclude, the present study reveals that vocal self-perception plays a role in the emergence of emotional feelings, reflecting an inferential process integrating both the monitoring of physiological changes and the interpretation of contextual information. Our findings support theories that see perception and cognitive interpretation of bodily signals as central for the construction of idiosyncratic emotional experiences (Barrett, 2017; Damasio & Carvalho, 2013), and suggest that, beyond interoception, exteroception of self-originated signals also plays a substantial role.

CRediT authorship contribution statement

Louise Goupil: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Petter Johansson:** Conceptualization, Writing - review & editing, Supervision. **Lars Hall Johansson:** Conceptualization, Writing - review & editing, Supervision. **Jean-Julien Aucouturier:** Conceptualization, Methodology, Software, Resources, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

The authors thank Louise Vasa and Lou Seropian for their help with collecting the data, as well as Marco Liuni, Pablo Arias, and Laura Rachman for their input regarding the voice transformation algorithm. Experimental data were collected and ethical approval obtained from the INSEAD/ Sorbonne University Center for Behavioural Science. This work was funded by ERC StG CREAM 335536 and ANR 2017 REFLETS (to J.J.A.), a H2020 Marie Skłodowska-Curie actions post-doctoral grant (845859 – JDIL, to L.G.), and a Swedish Research Council grant (2014-137, to PJ and LH).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2020.103072.

References

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., … David, A. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *ELife, 5*, 246–253.

Aucouturier, J.-J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences, 113*(4), 948–953.

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences, 110*(7), 2653–2658.

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology, 17*, 7–14.

Crivelli, C., & Fridlund, A. J. (2018). Facial Displays Are Tools for Social Influence. *Trends in Cognitive Sciences, 22*(5), 388–399.

Damasio, A. (2003). Feelings of emotion and the self. *Annals of the New York Academy of Sciences, 1001*(1), 253–261.

Damasio, & Carvalho, G. B. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews Neuroscience, 14*(2), 143–152.

Dezecache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. *Journal of Pragmatics, 59*(B), 221–233.

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour, 63*(3), 407–418.

Gaillard, R., Gourion, D., & Llorca, P. M. (2013). L'anhédonie dans la dépression. *L'Encéphale, 39*(4), 296–305.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Policy Analysis: Cambridge University Press.

Goupil, L., & Aucouturier, J.-J. (n.d.). Event-related prosody reveals distinct acoustic manifestations of accuracy and confidence in speech. https://doi.org/10.31234/OSF.IO/TUYNP.

Gray, M. A., Harrison, N. A., Wiens, S., & Critchley, H. D. (2007). Modulation of emotional appraisal by false physiological feedback during fMRI. *PLoS ONE, 2*(6), Article e546.

Goupil, L., Ponsot, E., Richardson, D., Reyes, G., Aucouturier, J. J. (2021). Listeners' perceptions of certainty and honesty of another speaker are associated with a common prosodic signature. Nature Communications.

Guerrier, G., Lellouch, L., Liuni, M., Vaglio, A., Rothschild, P. R., Baillard, C., & Aucouturier, J. J. (2019). Vocal markers of preoperative anxiety: a pilot study. *British Journal of Anaesthesia. Elsevier Ltd.*. https://doi.org/10.1016/j.bja.2019.06.020.

Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *The Journal of the Acoustical Society of America, 123*(1), 273–278. https://doi.org/10.1121/1.2817357.

James, W. (1890). *The Principles of Psychology Vol. 2. New York Holt* (Vol. 1). New York: Dover Publications.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication, 88*, 106–126.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America, 108*(3), 1246.

Kuznetsova, A., Brockhoff, P. B., & Christensen, H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R.

Laird, J. D., & Lacasse, K. (2014). Bodily influences on emotional feelings: Accumulating evidence and extensions of William James's theory of emotion. *Emotion Review, 6*(1), 27–34.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences, 114*(10), 2016–2025.

Max, L., Wauacet, M. E., & Vincene, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1053–1056).

Noah, T., Schul, Y., & Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of Personality and Social Psychology, 114*(5), 657–664.

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science, 2*(3), 281–296.

Park, H.-D., & Blanke, O. (2019). Coupling inner and outer body for self-consciousness. *Trends in Cognitive Sciences, 23*(5), 377–388.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.

Ponsot, E., Arias, P., & Aucouturier, J.-J. (2018). Uncovering mental representations of smiled speech using reverse correlation. The Journal of the Acoustical Society of America, 143(1), EL19–EL24.

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. *Acustica, 86*(6), 1038–1045.

Rachman, L., Liuni, M., Arias, P., Lind, A., Johansson, P., Hall, L., … Aucouturier, J. J. (2018). DAVID: An open-source platform for real-time transformation of infra-segmental emotional cues in running speech. *Behavior Research Methods, 50*(1), 323–343.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*(5), 379–399.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1–2), 227–256.

Scherer, S., Stratou, G., Gratch, J., & Morency, L.-P. (2013). Investigating voice quality as a speaker-independent indicator of depression and PTSD. Undefined.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology, 54*(5), 768–777.

Tian, X., & Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience, 27*(2), 352–364.

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., … Zwaan, R. A. (2016). Registered replication report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science, 11*(6), 917–928.

Wharton, T. (2009). *Pragmatics and non-verbal communication*. Cambridge: Cambridge University Press.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science, 24*(6), 947–956.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3*, 1212–1217.

Zimmermann, G., Quartier, V., Bernard, M., Salamin, V., & Maggiori, C. (2007). Qualités psychométriques de la version française de la TAS-20 et prévalence de l'alexithymie chez 264 adolescents tout-venant. *L'Encéphale, 33*(6), 941–946.