Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction

Jean-Julien Aucouturier, Petter Johansson, Lars Hall, Rodrigo Segnini, Lolita Mercadié, Katsumi Watanabe

Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction

Jean-Julien Aucouturier[a,1], Petter Johansson[b,c], Lars Hall[b], Rodrigo Segnini[d,e], Lolita Mercadié[f], and Katsumi Watanabe[g,h]

[a]Science and Technology of Music and Sound (STMS UMR9912), CNRS/Institut de Recherche et Coordination en Acoustique et Musique (IRCAM)/Université Pierre et Marie Curie (UPMC), 74005 Paris, France; [b]Lund University Cognitive Science, Lund University, 221 00 Lund, Sweden; [c]Swedish Collegium for Advanced Study, Uppsala University, 752 38 Uppsala, Sweden; [d]Siemens Healthcare, 141-8644 Tokyo, Japan; [e]Communication Science Laboratories, Nippon Telegraph and Telephone (NTT) Corporation, 243-0198 Kanagawa, Japan; [f]Laboratoire d'Etude de l'Apprentissage et du Développement (LEAD UMR5022), CNRS/Université de Bourgogne, 21000 Dijon, France; [g]Department of Intermedia Art and Science, Faculty of Science and Engineering, Waseda University, 169-8555 Tokyo, Japan; and [h]Research Center for Advanced Science and Technology, the University of Tokyo, 153 8904 Tokyo, Japan

Edited by Michael S. Gazzaniga, University of California, Santa Barbara, CA, and approved November 24, 2015 (received for review April 3, 2015)

Research has shown that people often exert control over their emotions. By modulating expressions, reappraising feelings, and redirecting attention, they can regulate their emotional experience. These findings have contributed to a blurring of the traditional boundaries between cognitive and emotional processes, and it has been suggested that emotional signals are produced in a goal-directed way and monitored for errors like other intentional actions. However, this interesting possibility has never been experimentally tested. To this end, we created a digital audio platform to covertly modify the emotional tone of participants' voices while they talked in the direction of happiness, sadness, or fear. The result showed that the audio transformations were being perceived as natural examples of the intended emotions, but the great majority of the participants, nevertheless, remained unaware that their own voices were being manipulated. This finding indicates that people are not continuously monitoring their own voice to make sure that it meets a predetermined emotional target. Instead, as a consequence of listening to their altered voices, the emotional state of the participants changed in congruence with the emotion portrayed, which was measured by both self-report and skin conductance level. This change is the first evidence, to our knowledge, of peripheral feedback effects on emotional experience in the auditory domain. As such, our result reinforces the wider framework of self-perception theory: that we often use the same inferential strategies to understand ourselves as those that we use to understand others.

emotion monitoring | vocal feedback | self-perception | digital audio effects | voice emotion

Over the last few years, tens of thousands of research articles have been published on the topic of emotion regulation, detailing how people try to manage and control emotion and how they labor to suppress expressions, reappraise feelings, and redirect attention in the face of tempting stimuli (1, 2). This kind of blurring of the traditional (antagonistic) boundaries between emotional and cognitive processes has gained more and more influence in the behavioral and neural sciences (3, 4). For example, a recent overview of neuroimaging and electrophysiological studies shows a substantial overlap of error monitoring and emotional processes in the dorsal mediofrontal cortex, lateral prefrontal areas, and anterior insula (5, 6). A consequence of this emerging integrative view is that emotional states and signals should be monitored in the same way as other intentional actions. That is, we ought to be able to commit emotional errors, detect them, and correct them. This assumption is particularly clear in the emotion as interoceptive inference view by Seth (7), which posits a central role for the anterior insular cortex as a comparator that matches top-down predictions against bottom-up prediction errors. However, there is a great need for novel empirical evidence to evaluate the idea of emotional error control, and we are not aware of any experimental tests in this domain.

The best candidate domain for experimentally inducing emotional errors is vocal expression. Vocal signals differ from other types of emotional display in that, after leaving the vocal apparatus and before reentering the auditory system, they exist for a brief moment outside of the body's sensory circuits. In principle, it should be possible to "catch" a vocal signal in the air, alter its emotional tone, and feed it back to the speaker as if it had been originally spoken this way. Such a manipulation would resemble the paradigm of speech perturbation, in which acoustic properties, like fundamental frequency ($F0$), are altered in real time and relayed back to the speakers, who are often found to monitor and compensate for the manipulation in their subsequent speech production (8, 9). Thus, would participants detect and correct feedback of a different emotional tone than they actually produced? If so, this behavior would provide novel experimental evidence in support of a dissolution of the cognition emotion divide. If not, it would provide a unique opportunity to study the effects of peripheral emotional feedback. As hypothesized by James–Lange-type theories of emotion (10–12), participants might then come to believe that the emotional tone was their own and align their feelings with the manipulation.

To this end, we aimed to construct three different audio manipulations that, in real time, make natural-sounding changes to a speaker's voice in the direction of happiness, sadness, or fear.

Significance

We created a digital audio platform to covertly modify the emotional tone of participants' voices while they talked toward happiness, sadness, or fear. Independent listeners perceived the transformations as natural examples of emotional speech, but the participants remained unaware of the manipulation, indicating that we are not continuously monitoring our own emotional signals. Instead, as a consequence of listening to their altered voices, the emotional state of the participants changed in congruence with the emotion portrayed. This result is the first evidence, to our knowledge, of peripheral feedback on emotional experience in the auditory domain. This finding is of great significance, because the mechanisms behind the production of vocal emotion are virtually unknown.

Author contributions: J.-J.A., P.J., L.H., R.S., L.M., and K.W. designed research; J.-J.A., P.J., R.S., and L.M. performed research; J.-J.A., P.J., and L.H. analyzed data; and J.-J.A., P.J., and L.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: aucouturier@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1506552113/-/DCSupplemental.

The manipulations use digital audio processing algorithms to simulate acoustic characteristics that are known components of emotional vocalizations (13, 14).

The happy manipulation modifies the pitch of a speaker's voice using upshifting and inflection to make it sound more positive (Audio Files S1–S8); it modifies its dynamic range using compression to make it sound more confident and its spectral content using high-pass filtering to make it sound more aroused (Fig. 1 and compare Audio File S1 with Audio File S2 and compare Audio File S5 with Audio File S6). Similarly, the sad manipulation operates on pitch using downshifting and spectral energy using a low-pass filter and a formant shifter (compare Audio File S1 with Audio File S3 and compare Audio File S5 with Audio File S7). The afraid manipulation operates on pitch using both vibrato and inflection (compare Audio File S1 with Audio File S4 and compare Audio File S5 with Audio File S8). The manipulations were implemented using a programmable hardware platform, allowing a latency of only 15 ms. (A low-latency, open-source software version of the voice manipulation is made available with this work at cream.ircam.fr.)

First, using three independent groups of Japanese speakers, we determined in a forced choice test that the manipulations were indistinguishable from natural samples of emotional speech ($n = 18$). Second, we verified that the manipulated samples were correctly associated with the intended emotions, whether these were described with valence arousal scales ($n = 20$) or free verbal descriptions ($n = 39$) (*SI Text*). Third, to assure that the manipulations were similarly perceived by these experiments' target population, we used the free verbal descriptions to construct a set of French adjective scales and let 10 French speakers rate the emotional quality of processed vs. nonprocessed sample voices. The six adjectives used were found to factor into two principal components, best labeled along the dimensions of positivity (happy/optimistic/sad) and tension (unsettled/anxious/relaxed). The three manipulations were perceived as intended: happy increased positivity and decreased tension, sad decreased positivity but did not affect tension, and afraid decreased positivity and increased tension (Fig. 2).

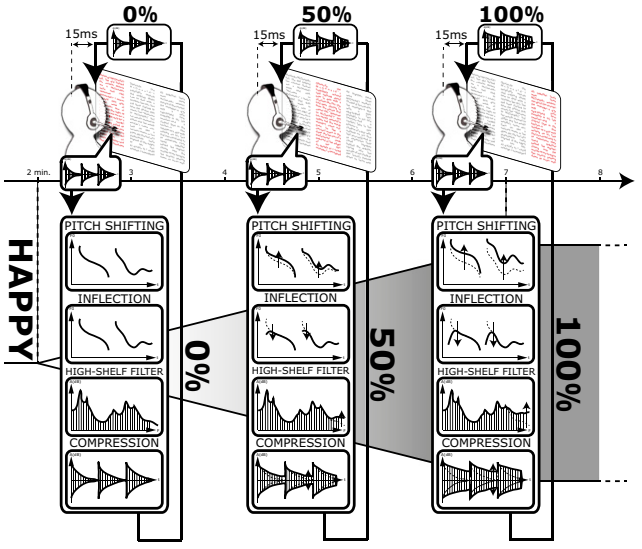

Fig. 1. Participants listened to themselves while reading, and the emotional tones of their voices were surreptitiously altered in the direction of happiness, sadness, or fear. In the happy condition (shown here), the speaker's voice is made to sound energized and positive using subtle variations of pitch (pitch-shifting and inflection), dynamic range (compression), and spectral energy (high-pass filter). The changes are introduced gradually from $t = 2$ to $t = 7$ min, and the feedback latency is kept constant across conditions at 15 ms. Example audio clips recorded in the experiment are available in Audio Files S5–S8.

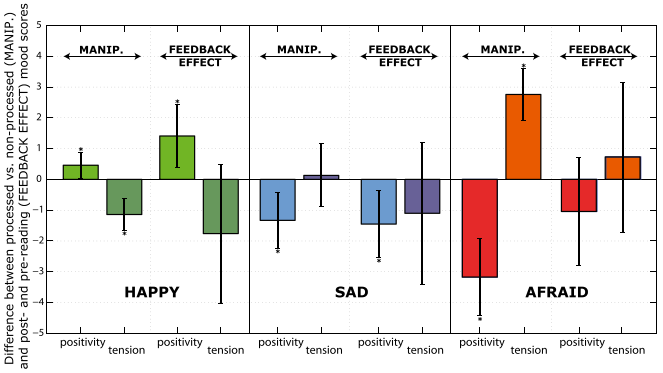

Fig. 2. Perceived difference in positivity and tension between processed and nonprocessed speech as judged by independent listeners and the post- and prereading changes in positivity and tension in the feedback experiment. Participants reading under manipulated feedback reported emotional changes consistent with the emotional characteristics of the voices that they heard. Error bars represent 95% confidence levels on the mean. The continuous scale is transformed to increments of 1 from −10 to +10. *Significant differences from zero at $P < 0.05$.

To determine whether participants would detect the induced emotional errors and measure possible emotional feedback effects of voice, we let participants ($n = 112$; female: 92) read an excerpt from a short story by Haruki Murakami while hearing their own amplified voice through a noise-cancelling headset. In the neutral control condition, the participants simply read the story from beginning to end. In three experimental conditions, the audio manipulations were gradually applied to the speaker's voice after 2 min of reading; after 7 min, the participants were hearing their own voice with maximum manipulation strength (Fig. 1). In total, the excerpt took about 12 min to read. The participants were asked to evaluate their emotional state both before and after reading using the same two-factor adjective scales previously used to classify the effects. In addition, we monitored the participants' autonomic nervous system responses while reading with their tonic skin conductance level (SCL). The participants were then asked a series of increasingly specific questions about their impression of the experiment to determine whether they had consciously detected the manipulation of their voice.

Results

Emotion Monitoring. Participant responses to posttest detection interviews were recorded by audio and written notes and then analyzed by the experimenters to categorize each participant into different detection levels. Only 1 participant (female; condition: afraid) reported complete detection ("you manipulated my voice to make it sound emotional"), and only 15 (female: 14; happy: 7, sad: 3, and afraid: 5) reported partial detection ("you did something to my voice; it sounded strange and it wasn't just the microphone"). The remaining 93 participants (female: 74; happy: 20, sad: 25, afraid: 21, and control: 27) reported no detection. To not bias any potential feedback results, the detecting participants were removed from all additional analyses. Three participants were also excluded because of technical problems with the feedback. The subsequent results, therefore, concern a total of 93 participants.

Feedback Effect. For the emotional self-rating task, the participants' scores on the previously described dimensions of positivity and tension were compared pre- and postreading. The scores (two levels: pre and post) were in significant interaction with the

type of manipulation (three levels): repeated-measure multivariate analysis of variance (rMANOVA) $F(4,124) = 3.30$, Wilk's $\Lambda = 0.81$, $P = 0.013$, $\alpha_{Bonferroni,2|4} = 0.016$. In the happy and sad conditions, the general pattern of the emotional changes matched how the manipulations were perceived in the pretest: happy feedback led to more positivity [$M = 7.4 > 6.9$; Fisher least-square difference (LSD), $P = 0.037$; Cohen's $d = 0.75$] but not significantly less tension ($M = 3.0 < 3.6$; Fisher LSD, $P = 0.14$); sad feedback led to less positivity ($M = 7.0 < 7.5$; Fisher LSD, $P = 0.017$; Cohen's $d = 0.70$) and as predicted, no significant change in tension ($M = 3.2 < 3.5$; Fisher LSD, $P = 0.29$). Despite being the most salient of the manipulations, we did not see significant emotional changes in the afraid condition for either positivity ($M = 6.5 < 6.8$; Fisher LSD, $P = 0.11$) or tension ($M = 3.8 < 4.0$; Fisher LSD, $P = 0.53$) (Fig. 2).

The evolution of participants' tonic SCL from minutes 3–8 was in significant interaction with the experimental condition [repeated-measure analysis of variance (rANOVA): $F(15,425) = 2.29$, $P = 0.0037$, $\alpha_{Bonferroni,1|4} = 0.0125$] (Fig. 3). SCL decreased the most in the control condition ($M = -5.9\%$ at $t = 8$) and less so in the sad ($M = -1.6\%$) and afraid conditions ($M = -2.4\%$), and it increased moderately in the happy condition ($M = +0.6\%$). SCLs reached at $t = 8$ were different from control in all three conditions (Fisher LSD, $P < 0.05$; Cohen's d: happy = 0.66, sad = 0.56, and afraid = 0.47). The steady decrease of tonic SCL seen in the control condition is the expected autonomic response associated with predictable and low-arousal control tasks, such as reading aloud (15). Although reports of systematic SCL dissociation between fear, sadness, and happiness are inconsistent (16), tonic SCL increase is typically associated with activated emotional states (17) as well as the appraisal of emotional speech or images (18, 19).

Audio Compensation. It is known that speakers reading under any kind of manipulated feedback may remain unaware of the audio alterations but compensate for them by, for example, adapting the pitch of their vocal production (8, 9). If such compensation occurred here, participants could be said to be monitoring their own expressive output, despite their lack of conscious awareness of the manipulation. Testing for such eventuality, we found no evidence of acoustical compensation in the participants' produced speech: the temporal evolution of the voices' fundamental frequencies, amplitudes, or voice qualities was not in significant statistical interaction with the experimental condition (*SI Text*). However, because participants were reading continuously a varied text of words as opposed to controlled phonemes as is often the case in pitch-altered feedback research, the variability in pitch over the course of speaking would make it difficult to detect compensation for the small pitch shifts used here (a 3-Hz increase in happy and a 3.5-Hz decrease in sad).

To further examine whether participants compensated for the emotional errors, even if they did not consciously detect them, we, therefore, replicated the first emotion-monitoring experiment with an additional emotional manipulation designed to feature drastically more pitch upshifting than before (+100 cents, a fourfold increase from happy) along with inflection and high-pass filtering. Applied to neutral speech, the resulting manipulation gave rise to a stressed, hurried impression (compare Audio File S9 with Audio File S10 and compare Audio File S11 with Audio File S12). Using the same adjective scales as above, we let 14 French speakers rate the emotional quality of processed vs. nonprocessed sample voices and found that this tensed manipulation differed significantly from neutral [multivariate $T^2 = 23.7$, $F(2,11) = 10.9$, $P = 0.0024$], with increased tension but no change of positivity.

Using this new manipulation, we then let $n = 90$ (all female) participants take part in a second emotion-monitoring experiment (neutral: 39, tensed: 38, and technical problems: 13). Results replicated both the low level of conscious detection and the emotional feedback found in experiment 1. First, only 2 of 38 tensed participants reported complete detection (5.6%), and 9 (23.6%) reported partial detection, proportions that did not differ from those in experiment 1. Second, scores of the nondetecting participants (tensed: 27 and control: 39) on the previously described dimensions of positivity and tension were compared pre- and postreading. The scores (two levels: pre and post) were in significant interaction with the condition [two levels; rMANOVA $F(2,42) = 4.10$, Wilk's $\Lambda = 0.83$, $P = 0.023$, $\alpha_{Bonferroni,1|2} = 0.025$] in a direction congruent with the new manipulation: more tension [$t(43) = 2.43$, $P = 0.019$; Cohen's $d = 0.70$] and no change of positivity [$t(43) = -1.94$, $P = 0.06$]. There was no interaction of the evolution of SCL with condition [rANOVA: $F(6,258) = 1.17$, $P = 0.32$, $\alpha_{Bonferroni,2|2} = 0.05$].

We extracted phonetical characteristics (mean $F0$, jitter, shimmer, and breathiness) from the manipulated (what's heard) and nonmanipulated (what's said) speech of nondetecting participants over successive 1-min windows from $t = 3$ to $t = 9$. First, we compared the manipulated and nonmanipulated speech of the tensed group and found that all four characteristics differed in line with the manipulation made, with increased pitch [$F(6,354) = 5.88$, $P = 0.0000$; +43 cents] and shimmer [$F(6,354) = 4.4$, $P = 0.0003$; +0.6%] and decreased jitter [$F(6,354) = 8.9$, $P = 0.0000$; −15 Hz] and breathiness [$F(6,354) = 8.3$, $P = 0.0000$; −0.3 dB]. This result shows that our method of analysis is sensitive enough to detect possible compensatory changes in voice production at least at a magnitude similar to that of the perturbation applied here. Second, we compared the nonmanipulated speech in the tensed group with the speech in the control group and found that the evolution of all four characteristics did not differ with condition. Thus, we again found no evidence that the participants compensated or otherwise adapted to the alterations (*SI Text* and Fig. S1).

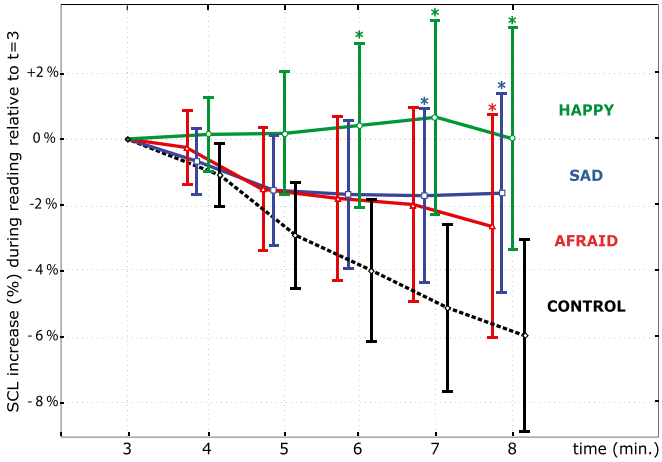

Fig. 3. Percentage increase of SCL over time measured relative to the level at the outset of manipulation (minute 3). Manipulation strength was gradually increased from 3 to 7 min and then, held at the highest level until the end of the task. Error bars represent 95% confidence intervals on the mean. *Time steps at which the SCL distribution is significantly different from the control condition (Fisher LSD, $P < 0.05$).

Discussion

In this study, we created a digital audio platform for real-time manipulation of the emotional tone of participants' voices in the direction of happiness, sadness, or fear. Classification results from both Japanese and French speakers revealed that the alterations were perceived as natural examples of emotional speech, corresponding to the intended emotions. This result

was robust across several different forms of rating formats. In experiment 1, the great majority of the participants remained unaware that their own voices were being transformed. As a consequence of listening to their altered voices, they came to react in congruence with the emotion portrayed as reflected in both self-report and skin conductance responses across the experiment. In experiment 2, we replicated key findings from experiment 1 and again, found no evidence that our participants vocally compensated for the altered audio feedback.

The low level of conscious detection of the manipulation as well as the absence of evidence of any compensation in the participants' production provide no support for the hypothesis that we continuously monitor our own voice to make sure that it meets a predetermined emotional target. This finding is significant because the neural processes underlying the production of emotional speech remain poorly understood (20, 21), and recent commentaries have suggested a central role of forward error-monitoring models in prosodic control (22–24). Our findings instead give support to dual-pathway models of vocal expression, where an older primate communication system responsible for affective vocalizations, like laughter and crying, penetrates the neocortex-based motor system of spoken language production, offering less opportunity for volitional control and monitoring than its cortical verbal counterpart (ref. 21, p. 542).

These results do not rule out the possibility that mismatch was registered below the threshold for conscious detection (25) and that the manipulated feedback overpowered any potential error signals (ref. 26 has a related discussion in the semantic domain). However, this suggestion would not explain why the nonconscious alarm was not acted on and especially, not compensated for in the participants' vocal productions. Similarly, it is interesting to speculate about the small minority of participants who actually detected the manipulation. If we assume a matrix of conflicting evidence in the task (from interoceptive signals and exteroceptive feedback), it is possible that their performance can be explain by individual differences in emotional sensitivity and awareness (27, 28).

When participants did not detect the manipulation, they instead attributed the vocal emotion as their own. This feedback result is as striking as the concomitant evidence for nondetection. The relationship between the expression and experience of emotions is a long-standing topic of heated disagreement in the field of psychology (10, 29, 30). Central to this debate, studies on facial feedback have shown that forced induction of a smile or a frown or temporary paralysis of facial muscles by botulinum injection leads to congruent changes in the participants' emotional reactions (11, 31–33). Although these experiments support the general notion that emotional expression influences experience, they all suffer from problems of experimental peculiarity and demand. Participants can never be unaware of the fact that they are asked to bite on a pencil to produce a smile or injected with a paralyzing neurotoxin in the eyebrows. In addition, these studies leave the causality of the feedback process largely unresolved: to what extent is it the (involuntary) production of an emotional expression or the afference from the expression itself that is responsible for feedback effects (33)? In contrast to all previous studies of feedback effects, we have created a situation where the participants produce a different signal than the feedback that they are receiving (in this case, neutral vs. happy, sad, afraid, or tensed). These conditions allow us to conclude that the feedback is the cause of the directional emotional change observed in our study. As such, our result reinforces the wider framework of self-perception theory: that we use our own actions to help infer our beliefs, preferences, and emotions (34, 35). Although we do not necessarily react the same way to emotion observed in ourselves and that observed in others, in both cases, we often use the same inferential strategies to arrive at our attributions (12, 36, 37).

In experiment 1, the happy and sad manipulations registered a feedback effect on the self-report measure but not the afraid voice, whereas all three manipulations differed from neutral on the SCL measure. It is unlikely that this outcome stemmed from different qualities of the manipulations, because all of them previously had been classified as the intended emotion (indeed, as can be seen in Fig. 2, the transformations to afraid separated most clearly from neutral in the discrimination test). Instead, we suggest to explain this unpredicted outcome by looking at the appraisal context of the experiment (38). Unlike previous studies, where the intensity of emotions was modulated by feedback, in our experiment, emotions were induced from scratch in relation to the same neutral backdrop in all conditions. However, most likely, the atmosphere of the short story that we used was more conducive to an emotional appraisal in terms of general mood changes, such as happy and sad (and later, tensed), compared with a more directional emotion, such as fear. In future studies, our aim will be to manipulate both context and feedback to determine the relative importance of each influence.

Alternatively, it should be noted that, although concordance between different measures, such as self-report and psychophysiology, is often posited by emotion theories, the empirical support for this position is not particularly strong (39, 40). Thus, a dual-systems view of emotion could, instead, interpret an effect on the SCL profile but not on self-report as unconscious emotional processing (25). This interpretation might be particularly fitting for an emotion like fear, where evidence indicates the existence of a unconscious subcortical route through which emotional stimuli quickly reach the amygdala (41).

In summary, this result gives novel support for modular accounts of emotion production and self-perception theory and argues against emotional output monitoring. In future experiments, we will tie our paradigm closer to particular models of speech production (42, 43) and explore the interesting discrepancies between our results and the compensation typically found in pitch perturbation studies. In addition, real-time emotional voice manipulation allows for a number of further paths of inquiry. For example, in the field of decision-making, emotion is often seen as integral to both rapid and deliberate choices (44), and it seems likely that stating preferences and choosing between options using emotionally altered speech might function as somatic markers (45) and influence future choices. More speculatively, emotion transformation might have remedial uses. It has been estimated that 40–75% of all psychiatric disorders are characterized by problems with emotion regulation (46). Thus, it is possible that positive attitude change can be induced from retelling of affective memories or by redescribing emotionally laden stimuli and events in a modified tone of voice. Finally, outside academia, we envisage that our paradigm could be used to enhance the emotionality of live singing performances as well as increase immersion and atmosphere in online gaming, where vocal interactions between players often lack an appropriate emotional edge.

Materials and Methods

Experiment 1: Audio Manipulations. The happy effect processed the voice with pitch-shifting, inflection, compression, and a high shelf filter (definitions are in *SI Text*). Pitch-shifting was set to a positive shift of +25 cents. Inflection had an initial pitch shift of −50 cents and a duration of 400 ms. Compression had a −26-dB threshold, 4:1 soft-knee ratio, and 10 dB/s attack and release. High shelf-filtering had a shelf frequency of 8,000 Hz and a high-band gain of 10 dB per octave. The sad effect processed the voice with pitch-shifting, a low shelf filter, and a formant shifter. Pitch-shifting had a negative shift of −30 cents. Low shelf-filtering had a cutoff frequency 8,000 Hz and a high-band roll off of 10 dB per octave. Formant shifting used a tract ratio of 0.9. Finally, the afraid effect processed the voice with vibrato and inflection. Vibrato was sinusoidal with a depth of 15 cents and frequency of 8.5 Hz. Inflection had an initial pitch shift of +120 cents and a duration of 150 ms. The effects were implemented

with a programmable hardware platform (VoicePro, TC-Helicon; TC Group Americas) with an in/out latency of exactly 15 ms.

Pilot experiment. A sentence from the French translation of the short story collection *The Elephant Vanishes* by Haruki Murakami was recorded in a neutral tone by eight (male: four) relatively young ($M = 20.1$) native French speakers. Recordings were processed by each of the audio effects (happy, sad, and afraid), resulting in 24 different pairs of one neutral reference and one processed variant thereof (eight trials per effect). We then asked $n = 10$ independent listeners (male: five) from the same population to judge the emotional content of the processed voices compared with their neutral reference using six continuous scales anchored with emotional adjectives (happy, optimistic, relaxed, sad, anxious, and unsettled). For analysis, response data were factored into two principal components (with varimax rotation; 91% total variance explained), with factors suggesting labels of positivity (happy, optimistic, and sad: 80% variance explained) and tension (unsettled, anxious, and relaxed: 11% variance explained). The manipulations were perceived to be distinct from one another on both dimensions [multivariate $F(4,6) = 8.33$, $P = 0.013$]. Emotional ratings of manipulated speech differed from nonmanipulated speech for happy [multivariate $T^2 = 28.6$, $F(2,8) = 12.7$, $P = 0.003$] with increased positivity [$t(9) = 2.51$, $P = 0.03$; Cohen's $d = 1.67$] and decreased tension [$t(9) = -4.98$, $P = 0.0008$; Cohen's $d = 3.32$], sad [multivariate $T^2 = 11.3$, $F(2,8) = 5.0$, $P = 0.038$] with decreased positivity [$t(9) = -3.34$, $P = 0.008$; Cohen's $d = 2.22$] and unchanged tension [$t(9) = 0.30$, $P = 0.77$], and afraid [multivariate $T^2 = 54.3$, $F(2,8) = 24.1$, $P = 0.0004$] with decreased positivity [$t(9) = -5.7$, $P = 0.0003$; Cohen's $d = 3.8$] and increased tension [$t(9) = 7.34$, $P = 0.00004$; Cohen's $d = 4.8$] (Fig. 2).

Feedback procedure. Participants were recruited to perform two successive Stroop tasks separated by the main reading task, which was presented as a filler task. At the beginning of the experiment, participants were fitted with two finger electrodes (Biosemi BioPaC MP150) on their nondominant hands, from which their continuous SCLs were measured throughout the session. After the first Stroop task, participants were asked to evaluate their emotional state using six continuous adjective scales. For the reading task, participants were fitted with noise-cancelling headsets (Sennheiser HME-100) with attached microphones, in which they could hear their own amplified voices while they read out loud. They were tasked to read an excerpt from a short story collection by Haruki Murakami ("The Second Bakery Attack" from *The Elephant Vanishes*), and text was comfortably presented on a board facing them. In the neutral control condition, the participants simply read the story from beginning to end. In three experimental conditions, the emotional effects were gradually applied to the speaker's voice after 2 min of reading. The strength of the effects increased by successive increments of their parameter values triggered every 2 min by messages sent to the audio processor from an audio sequencer (Steinberg Cubase 7.4). Levels were calibrated using a Bruël & Kjær 2238 Mediator Sound-Level Meter (Bruël & Kjær Sound & Vibration), and overall effect gain was automatically adjusted so that gradual increases in effect strength did not result in gradual increases of sound level. After 7 min, the participants were hearing their own voices with the maximum of the effect added until the end of the reading task. After the reading task and before the second Stroop task, participants were again asked to evaluate their emotional state using adjective scales. In addition, they were also asked to fill in the Profile of Mood States (POMS) questionnaire and evaluate the emotional content of the text using a brief Self-Assessment Manikin (SAM) test (results for the POMS, the SAM, and Stroop are not discussed in the text) (*SI Text*). After the second Stroop task, participants were then asked a series of increasingly specific questions about their impressions of the experiment to determine whether they had consciously detected the manipulations of their voices. Finally, participants were debriefed and informed of the true purpose of the experiment.

Participants. In total, $n = 112$ (female: 92) participants took part in the study, and all were relatively young ($M = 20.1$, $SD = 1.9$) French psychology undergraduates at the University of Burgundy in Dijon, France. The students were rewarded for their participation by course credits. Three participants were excluded who could not complete the feedback part of the experiment because of technical problems, leaving $n = 109$ (female: 89) for subsequent analysis.

Detection questionnaire. At the end of the experiment, participants were asked a series of increasingly specific questions to determine whether they had consciously detected the manipulations of their voices. Participants were asked (*i*) what they had thought about the experiment, (*ii*) whether they had noticed anything strange or unusual about the reading task, (*iii*) whether they had noticed anything strange or unusual about the sound of their voices during the reading task, and (*iv*) because a lot of people do not like to hear their own voices in a microphone, whether that was what they meant by unusual in this case. Answers to all questions were recorded by audio and written notes and then, analyzed by the experimenters to categorize each participant into four detection levels: (*i*) "you manipulated my voice to make it sound emotional" (complete detection), (*ii*) "you did something to my voice; it sounded strange and it was not just the microphone or headphones" (partial detection), (*iii*) "my voice sounded unusual, and I am confident that it was because I was hearing myself through headphones" (no detection), and (*iv*) "there was nothing unusual about my voice" (no detection).

Skin conductance. The participants' SCLs were continuously recorded during the complete duration of the reading. Data were acquired with gain of 5 micro-ohm/volt, sampled at 200 Hz, and low pass-filtered with a 1-Hz cutoff frequency. SCLs were averaged over nonoverlapping 1-min windows from $t = 3$ min to $t = 8$ min and normalized relative to the level at $t = 3$.

Mood scales. Feedback participants reported their emotional states both before and after the reading task using the same six adjective scales used in pilot data. Responses were combined into positivity (happy, optimistic, and sad) and tension (unsettled, anxious, and relaxed) averaged scores, and their differences were computed pre- and postreading.

Correction for multiple measures. Results for scales and SCLs were corrected for multiple measures (four measures of emotional feedback: scales, SCL, the POMS, and Stroop) using Holm's sequential Bonferroni procedure. The two measures of error detection (detection rate and audio compensation) were not corrected, because detection rate is a descriptive measure.

Experiment 2: Audio Manipulation. The tensed manipulation consisted of pitch-shifting (+100 cents; a fourfold increase from happy), inflection (initial pitch shift of +150 cents and a duration of 150 ms), and high shelf-filtering (shelf frequency of 8,000 Hz; +10 dB per octave). The effect was implemented with a software platform based on the Max/MSP language designed to reproduce the capacities of the hardware used in experiment 1, and it is available at cream.ircam.fr.

Pilot experiment. Eight recordings of the same sentence spoken in a neutral tone by eight young female native French speakers were processed with the tensed effect and presented paired with their nonmanipulated neutral reference to $n = 14$ French speakers (male: 5) who rated their emotional quality using the same adjective scales used in the main experiment. Participants found that tensed manipulated speech differed from nonmanipulated speech [multivariate $T^2 = 23.7$, $F(2,11) = 10.9$, $P = 0.002$] and with increased tension [$M = +4.4$, $t(13) = 3.39$, $P = 0.005$; Cohen's $d = 1.88$] but found no change of positivity [$M = +0.1$, $t(13) = 0.08$, $P = 0.93$].

Feedback procedure. The same procedure as in experiment 1 was used, with the same text read under one manipulated (tensed) condition and one control condition. Measures were the same, with the exception of the POMS, the SAM, and Stroop tasks, which were not used in experiment 2.

Participants. Ninety (all female) participants took part in the study; all were relatively young ($M = 21.0$, $SD = 2.3$) undergraduate students at Sorbonne University (Paris, France). Participants were rewarded for their participation by cash; 13 participants were excluded who could not complete the feedback task because of technical problems, leaving 77 participants (neutral: 39 and tensed: 38).

Correction for multiple measures. Results for scales and SCLs were corrected for multiple measures (two measures of emotional feedback: scales and SCL) using Holm's sequential Bonferroni procedure. The two results of error detection (detection rate and audio compensation) were not corrected for on multiple measures, because one of them, the detection rate, is not used in any statistical tests.

The procedures used in this work were approved by the Institutional Review Boards of the University of Tokyo, of the INSERM, and of the Institut Européen d'Administration des Affaires (INSEAD). In accordance with the American Psychological Association Ethical Guidelines, all participants gave their informed consent and were debriefed and informed about the true purpose of the research immediately after the experiment.

ACKNOWLEDGMENTS. J.-J.A. acknowledges the assistance of M. Liuni and L. Rachman [Institut de Recherche et Coordination en Acoustique et Musique (IRCAM)], who developed the software used in experiment 2, and H. Trad (IRCAM), who helped with data collection. Data in experiment 2 was collected at the Centre Multidisciplinaire des Sciences Comportementales Sorbonne Universités–Institut Européen d'Administration des Affaires (INSEAD). All data reported in the paper are available on request. The work was funded, in Japan, by two Postdoctoral Fellowships for Foreign Researchers to the Japanese Society for the Promotion of Science (JSPS; to J.-J.A. and P.J.), the Japanese Science and Technology (JST) ERATO Implicit Brain Function Project (R.S. and K.W.), and a JST CREST Project (K.W.). Work in France was partly funded by European Research Council Grant StG-335536 CREAM (to J.-J.A.) and the Foundation of

the Association de Prévoyance Interprofessionnelle des Cadres et Ingénieurs de la région Lyonnaise (APICIL; L.M.). In Sweden, P.J. was supported by the Bank of Sweden Tercentenary Foundation and Swedish Research Council Grant 2014-1371, and L.H. was supported by Bank of Sweden Tercentenary Foundation Grant P13-1059:1 and Swedish Research Council Grant 2011-1795.

1. Gross JJ (2015) Emotion regulation?: Current status and future prospects. *Psychol Inq* 26(1):1–26.
2. Moyal N, Henik A, Anholt GE (2014) Cognitive strategies to regulate emotions-current evidence and future directions. *Front Psychol* 4(2014):1019.
3. Keltner D, Horberg EJ (2015) Emotion cognition interactions. *APA Handbook of Personality and Social Psychology*, APA Handbooks in Psychology, eds Mikulincer M, Shaver PR (American Psychological Association, Washington, DC), Vol 1, pp 623–664.
4. Inzlicht M, Bartholow BD, Hirsh JB (2015) Emotional foundations of cognitive control. *Trends Cogn Sci* 19(3):126–132.
5. Koban L, Pourtois G (2014) Brain systems underlying the affective and social monitoring of actions: An integrative review. *Neurosci Biobehav Rev* 46(Pt 1):71–84.
6. Cromheeke S, Mueller SC (2014) Probing emotional influences on cognitive control: An ALE meta-analysis of cognition emotion interactions. *Brain Struct Funct* 219(3): 995–1008.
7. Seth AK (2013) Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci* 17(11):565–573.
8. Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103(6):3153–3161.
9. Jones JA, Munhall KG (2000) Perceptual calibration of F0 production: Evidence from feedback perturbation. *J Acoust Soc Am* 108(3 Pt 1):1246–1251.
10. James W (1890) *Principles of Psychology* (Holt, New York), Vol 2.
11. Flack W (2006) Peripheral feedback effects of facial expressions, bodily postures, and vocal expressions on emotional feelings. *Cogn Emotion* 20(2):177–195.
12. Laird JD, Lacasse K (2013) Bodily influences on emotional feelings: Accumulating evidence and extensions of William James's theory of emotion. *Emot Rev* 6(1):27–34.
13. Briefer EF (2012) Vocal expression of emotions in mammals: Mechanisms of production and evidence. *J Zool* 288(1):1–20.
14. Juslin P, Scherer K (2005) Vocal expression of affect. *The New Handbook of Methods in Nonverbal Behavior Research*, eds Harrigan J, Rosenthal R, Scherer K (Oxford Univ Press, Oxford), pp 65–135.
15. Nagai Y, Critchley HD, Featherstone E, Trimble MR, Dolan RJ (2004) Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: A physiological account of a "default mode" of brain function. *Neuroimage* 22(1): 243–251.
16. Kreibig SD (2010) Autonomic nervous system activity in emotion: A review. *Biol Psychol* 84(3):394–421.
17. Silvestrini N, Gendolla GH (2007) Mood effects on autonomic activity in mood regulation. *Psychophysiology* 44(4):650–659.
18. Aue T, Cuny C, Sander D, Grandjean D (2011) Peripheral responses to attended and unattended angry prosody: A dichotic listening paradigm. *Psychophysiology* 48(3): 385–392.
19. Lang PJ, Greenwald MK, Bradley MM, Hamm AO (1993) Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30(3):261–273.
20. Pichon S, Kell CA (2013) Affective and sensorimotor components of emotional prosody generation. *J Neurosci* 33(4):1640–1650.
21. Ackermann H, Hage SR, Ziegler W (2014) Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav Brain Sci* 37(6):529–546.
22. Frühholz S, Sander D, Grandjean D (2014) Functional neuroimaging of human vocalizations and affective speech. *Behav Brain Sci* 37(6):554–555.
23. Hasson U, Llano DA, Miceli G, Dick AS (2014) Does it talk the talk? On the role of basal ganglia in emotive speech processing. *Behav Brain Sci* 37(6):556–557.
24. Pezzulo G, Barca L, D'Ausilio A (2014) The sensorimotor and social sides of the architecture of speech. *Behav Brain Sci* 37(6):569–570.
25. Gainotti G (2012) Unconscious processing of emotions and the right hemisphere. *Neuropsychologia* 50(2):205–218.
26. Lind A, Hall L, Breidegard B, Balkenius C, Johansson P (2014) Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychol Sci* 25(6):1198–1205.
27. Garfinkel SN, Seth AK, Barrett AB, Suzuki K, Critchley HD (2015) Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biol Psychol* 104:65–74.
28. Kuehn E, Mueller K, Lohmann G, Schuetz-Bosbach S (January 23, 2015) Interoceptive awareness changes the posterior insula functional connectivity profile. *Brain Struct Funct*.
29. Darwin C (1872) *The Expression of Emotions in Man and Animals* (Philosophical Library, New York).
30. Schachter S, Singer JE (1962) Cognitive, social, and physiological determinants of emotional state. *Psychol Rev* 69:379–399.
31. Strack F, Martin LL, Stepper S (1988) Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *J Pers Soc Psychol* 54(5):768–777.
32. Havas DA, Glenberg AM, Gutowski KA, Lucarelli MJ, Davidson RJ (2010) Cosmetic use of botulinum toxin-a affects processing of emotional language. *Psychol Sci* 21(7): 895–900.
33. Hennenlotter A, et al. (2009) The link between facial feedback and neural activity within central circuitries of emotion–new insights from botulinum toxin-induced denervation of frown muscles. *Cereb Cortex* 19(3):537–542.
34. Dennett DC (1987) *The Intentional Stance* (MIT Press, Cambridge, MA).
35. Johansson P, Hall L, Tarning B, Sikstrom S, Chater N (2014) Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it! *J Behav Decis Making* 27(3):281–289.
36. Bem DJ (1972) Self-perception theory. *Advances in Experimental Social Psychology*, ed Berkowitz L (Academic, New York), Vol 6, pp 1–62.
37. Laird JD (1974) Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *J Pers Soc Psychol* 29(4):475–486.
38. Gray MA, Harrison NA, Wiens S, Critchley HD (2007) Modulation of emotional appraisal by false physiological feedback during fMRI. *PLoS One* 2(6):e546.
39. Hollenstein T, Lanteigne D (2014) Models and methods of emotional concordance. *Biol Psychol* 98:1–5.
40. Evers C, et al. (2014) Emotion response coherence: A dual-process perspective. *Biol Psychol* 98:43–49.
41. Adolphs R (2013) The biology of fear. *Curr Biol* 23(2):R79–R93.
42. Hickok G (2012) Computational neuroanatomy of speech production. *Nat Rev Neurosci* 13(2):135–145.
43. Pickering MJ, Garrod S (2013) An integrated theory of language production and comprehension. *Behav Brain Sci* 36(4):329–347.
44. Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. *Annu Rev Psychol* 66:799–823.
45. Batson CD, Engel CL, Fridell SR (1999) Value judgments: Testing the somatic-marker hypothesis using false physiological feedback. *Pers Soc Psychol Bull* 25(8):1021–1032.
46. Gross JJ, Jazaieri H (2014) Emotion, emotion regulation, and psychopathology: An affective science perspective. *Clin Psychol Sci* 2(4):387–401.
47. Lang P (1980) *Behavioural Treatment and Bio-Behavioural Assessment: Computer Applications* (Ablex, Norwood, NJ), pp 119–137.
48. McNair D, Lorr M, Droppleman L (1971) *Profile of Mood States (POMS) Manual* (Educational and Industrial Testing Service, San Diego).
49. Williams JMG, Mathews A, MacLeod C (1996) The emotional Stroop task and psychopathology. *Psychol Bull* 120(1):3–24.
50. Bonin P, et al. (2003) Normes de concrétude de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *Annee Psychol* 103(4):655–694.