Cracking the social code of speech prosody using reverse correlation

Emmanuel Ponsot, Juan José Burred, Pascal Belin, Jean-Julien Aucouturier

▶ **To cite this version:**

Emmanuel Ponsot, Juan José Burred, Pascal Belin, Jean-Julien Aucouturier. Cracking the social code of speech prosody using reverse correlation. Proceedings of the National Academy of Sciences of the United States of America, 2018, 115 (15), pp.3972-3977. 10.1073/pnas.1716090115 . hal-02004519

HAL Id: hal-02004519
https://hal.science/hal-02004519

Submitted on 19 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cracking the social code of speech prosody using reverse correlation

Emmanuel Ponsot[a,b,c,1], Juan José Burred[d], Pascal Belin[e,f,g,2], and Jean-Julien Aucouturier[a,h,2]

[a]Sciences et Technologies de la Musique et du Son, UMR 9912, Institut de Recherche et Coordination Acoustique/Musique, CNRS and Sorbonne Universitè, 75004 Paris, France; [b]Laboratoire des Systèmes Perceptifs, UMR 8248, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France; [c]Département d'Études Cognitives, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France; [d]Independent Researcher, 75013 Paris, France; [e]Institut de Neurosciences de la Timone, UMR 7289, Centre National de la Recherche Scientifique and Aix-Marseille Université, 13007 Marseille, France; [f]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QQ, United Kingdom; [g]Département de Psychologie, Université de Montréal, Montréal, QC H3T 1J4, Canada; and [h]Okanoya Emotional Information Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Wako, Saitama 332-0012, Japan

Edited by Dale Purves, Duke University, Durham, NC, and approved February 12, 2018 (received for review September 12, 2017)

Human listeners excel at forming high-level social representations about each other, even from the briefest of utterances. In particular, pitch is widely recognized as the auditory dimension that conveys most of the information about a speaker's traits, emotional states, and attitudes. While past research has primarily looked at the influence of mean pitch, almost nothing is known about how intonation patterns, i.e., finely tuned pitch trajectories around the mean, may determine social judgments in speech. Here, we introduce an experimental paradigm that combines state-of-the-art voice transformation algorithms with psychophysical reverse correlation and show that two of the most important dimensions of social judgments, a speaker's perceived dominance and trustworthiness, are driven by robust and distinguishing pitch trajectories in short utterances like the word "Hello," which remained remarkably stable whether male or female listeners judged male or female speakers. These findings reveal a unique communicative adaptation that enables listeners to infer social traits regardless of speakers' physical characteristics, such as sex and mean pitch. By characterizing how any given individual's mental representations may differ from this generic code, the method introduced here opens avenues to explore dysprosody and social-cognitive deficits in disorders like autism spectrum and schizophrenia. In addition, once derived experimentally, these prototypes can be applied to novel utterances, thus providing a principled way to modulate personality impressions in arbitrary speech signals.

speech | voice | prosody | social traits | reverse-correlation

In social encounters with strangers, human beings are able to form high-level social representations from very thin slices of expressive behavior (1) and quickly determine whether the other is a friend or a foe and whether they have the ability to enact their good or bad intentions (2, 3). While much is already known about how facial features contribute to such evaluations (4, 5), determinants of social judgments in the auditory modality remain poorly understood. Even when we cannot see others, we can instantly process their voice to infer, e.g., whether they can be trusted (6). In particular, voice height, or pitch, is widely recognized as the auditory dimension that conveys most of the information about a speaker's traits or states, not only by virtue of its mean value (7–9), but also, and perhaps primarily, by its intonation, i.e., its temporal pattern of variation around the mean.

Anthropologists, linguists, and psychologists have noted regularities of pitch contours in social speech for decades. Notably, patterns of high or rising pitch are associated with social traits such as submissiveness or lack of confidence, and low or falling pitch with dominance or self-confidence (10, 11), a code that has been proposed to be universal across species (12). Unfortunately, because these observations stem either from acoustic analysis of a limited number of actor-produced utterances or from the linguistic analysis of small ecological corpora, it has remained difficult to attest of their generality and causality in cognitive mechanisms, and we still do not know what exact pitch contour maximally elicits social percepts.

Inspired by a recent series of powerful data-driven studies in visual cognition in which facial prototypes of social traits were derived from human judgments of thousands of computer-generated visual stimuli (13–16), we developed a voice-processing algorithm able to manipulate the temporal pitch dynamics of arbitrary recorded voices in a way that is both fully parametric and realistic and used this technique to generate thousands of novel, natural-sounding variants of the same word utterance, each with a randomly manipulated pitch contour (Fig. 1 and *Materials and Methods*). We then asked human listeners to evaluate the social state of the speakers for each of these manipulated stimuli and reconstructed their mental representation of what speech prosody drives such judgments, using the psychophysical technique of reverse correlation.

The reverse-correlation technique presents a system (here, a human listener) with a slightly perturbed stimulus over many trials. This perturbation may be created by directly adding white noise on a stimulus, or, as we do here with pitch, by manipulating its higher-level dimensions by using random deviations around baseline (Fig. 1, *Left*). Perturbed stimuli will, on different trials, lead to different responses of the system, and the tools of

Significance

In speech, social evaluations of a speaker's dominance or trustworthiness are conveyed by distinguishing, but little-understood, pitch variations. This work describes how to combine state-of-the-art vocal pitch transformations with the psychophysical technique of reverse correlation and uses this methodology to uncover the prosodic prototypes that govern such social judgments in speech. This finding is of great significance, because the exact shape of these prototypes, and how they vary with sex, age, and culture, is virtually unknown, and because prototypes derived with the method can then be reapplied to arbitrary spoken utterances, thus providing a principled way to modulate personality impressions in speech.

Author contributions: E.P., P.B., and J.-J.A. designed research; E.P. performed research; E.P. and J.J.B. contributed new reagents/analytic tools; E.P. analyzed data; and E.P., P.B., and J.-J.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: All audio files, individual data, and analysis code are freely available at https://zenodo.org/record/1186278.

[1]To whom correspondence should be addressed. Email: emmanuel.ponsot@ens.fr.

[2]J.-J.A. and P.B. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716090115/-/DCSupplemental.

Published online March 26, 2018.

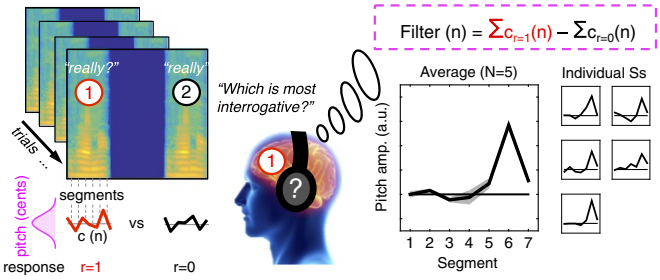

Fig. 1. Accessing mental representations of interrogative prosody by using reverse correlation. To validate the paradigm used in this study, we examined prosodic prototypes related to the evaluation of interrogative vs. declarative utterances. (*Left*) Utterances of the same word "vraiment" ("really") were digitally manipulated to have random pitch contours *c*(*n*).. Participants were presented pairs of manipulated words and judged which was most interrogative. (*Right*) Prosodic mental representations, or prototypes, were computed as the mean pitch contour of the voices perceived as interrogative ("really?"), minus those judged declarative ("really."). As predicted, the prototypes associated with interrogative judgments showed a clear pitch increase at the end of the second syllable, which was observable both in averaged and in individual prototypes. amp., amplitude.

reverse correlation can be used to infer the system's functional properties from the pattern of stimulus noise and their associated responses. In particular, first-order reverse-correlation analysis (as we do here) assumes a computational model in which participant evaluations are based on the distance of a given trial's noise field to an internal template, or prototype. In a two-interval task, this internal template can be estimated by using the so-called "classification image" technique (17), which simply consists of summing the noise fields that led to a positive answer and subtracting those that led to a negative answer (Fig. 1, *Right*). The technique was first used by psychophysicists to characterize human sensory processing (18), but it is also a powerful tool to characterize higher-level perceptual or cognitive processes, most notably in vision, for which it can uncover the "optimal stimulus" (or "mental representation") that is driving participants' responses (for a review, see, e.g., refs. 5 and 14).

To validate the approach, we conducted a preliminary experiment in which $n = 5$ observers had to categorize utterances as interrogative or declarative. We recorded a 426-ms utterance of the French word "vraiment" ("really"), and generated prosodic variations by dividing it into six segments of 71 ms and randomly manipulating the pitch of each breakpoint independently using Gaussian distributions (see Fig. 1, *Left* and *Materials and Methods* for details). We presented each participant with 700 pairs of such manipulated utterances ("really/really?"), asking them to judge which sounded more interrogative. As predicted, reverse-correlating observers' responses revealed mental representations of interrogative prosody showing a consistent marked increase of pitch at the end of the second syllable (Fig. 1, *Right*).

We then used the same paradigm to probe the mechanisms of pitch contour processing engaged in person perception in social encounters with strangers. A wealth of research in the past 20 y has shown that social evaluations in such situations are driven by two dimensions, of likability/trustworthiness/warmth and efficacy/dominance/competence (2, 3), which, in keeping with the recent literature on face (4) and voice perception (6), we labeled here as trustworthiness and dominance. Specifically, we considered here judgments of dominance and trustworthiness in spoken utterances of the French word "bonjour" ("hello") (Fig. S1). We presented two independent groups of $n = 21$ and $n = 23$ participants with hundreds of pairs of random pronunciations of that word, created either from the single recording of one male or one female speaker (see details in Fig. S2), and asked them to indicate which of the two variants in each pair they perceived as most dominant/trustworthy.

Results

Derivation of Dominance and Trustworthiness Prototypes. We first analyzed how participants' judgments varied with the mean pitch of the manipulated utterances. Consistent with previous studies, perceived dominance was negatively related to mean pitch (19, 20) and trustworthiness positively related to mean pitch (6, 8), although more weakly so than dominance (Fig. 2*B*). Beyond mean pitch, we then analyzed dynamic pitch contours with reverse correlation and found that dominance judgments were driven by pitch prototypes with a gradual pitch decrease on both syllables, while pitch prototypes for trustworthiness showed a rapid pitch increase on the second syllable only (Fig. 2*D*). The two patterns were decidedly dynamic in time: Statistical analyses showed significant effects of segment position in both tasks ($P < 0.001$), and the slopes of linear best fits for segments 2–5 were

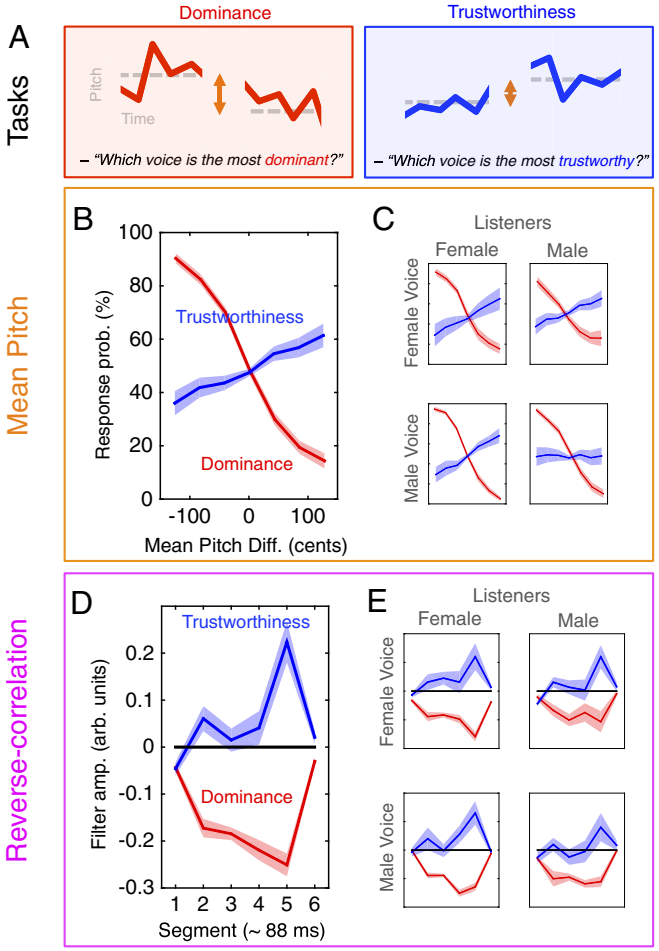

Fig. 2. Effects of mean pitch and pitch dynamics on judgments of social dominance and trustworthiness. (*A*) In each task (dominance or trustworthiness), participants compared two randomly modulated voices. (*B*) Response probability (prob.) for dominance (red) and trustworthiness (blue) as a function of mean pitch difference (diff.) in each pair: Lower voices in each pair were judged to be more dominant and, to a lesser extent, less trustworthy. (*C*) The influence of mean pitch is stable across stimuli and listener gender (see detailed analyses in *Supporting Information*). (*D*) Normalized mental prototypes of pitch contours (i.e., first-order reverse-correlation kernels) in the two tasks (*Materials and Methods*). (*E*) These prototypes were strikingly stable across stimuli and listener gender. Shaded areas, SEM. amp., amplitude; arb., arbitrary.

significantly negative for dominance and positive for trustworthiness (all $P < 0.05$).

Importantly, we found that male and female listeners judged dominance, and trustworthiness, in a similar fashion (illustrated in Fig. 2 *C* and *E*), with no effect of listeners' gender nor interactions with stimuli gender for either task (all $P > 0.05$). There was a small difference between the prototypes obtained for the male vs. the female voice for dominance, but this difference was only visible on a single time segment (fourth), likely explained by intrasyllabic loudness contour differences between the two voices (*Supporting Information*). Finally, while we observed interindividual differences regarding the exact shape of these patterns, especially for trustworthiness (Fig. S3), the robustness of the internal prototypes for each listener was evident. In particular, strikingly similar prototypes were obtained for male and female voices, even though these were measured for a given listener on different days.

Application to Novel Utterances. In a second experiment, we tested the generality of these prototypes across words and speakers by applying them, their opposite patterns, or their mean values to new recordings of "bonjour," as well as to a variety of other two-syllable words recorded by new speakers (*Materials and Methods* and Fig. S1). Two new groups of participants ($n = 21$ and $n = 19$) rated the perceived dominance and trustworthiness of these new transformed voices, randomly mixed in terms of content and speaker. As predicted for dominance, applying the original prototype to novel utterances significantly increased their perceived dominance, whereas applying the opposite pattern significantly decreased it (all $P < 0.001$; Fig. 3), both for "hello" and novel words [although more weakly so: Condition × WordType interaction; $F(5, 85) = 8.6$, $P < 0.001$, $\eta_p^2 = 0.34$, $\tilde{\varepsilon} = 0.81$]. Even though dominant prototypes flattened to their mean pitch value also led to a strong increase of perceived dominance, this increase was significantly smaller than for original prototypes, showing that the prototypes did not reduce to a simple mean pitch effect. Finally, applying the trustworthiness prototype and its opposite pattern significantly degraded and improved perceived dominance, respectively, but these effects were significantly smaller than those induced by the appropriate dominance prototypes ($P < 0.01$), showing that the two prototypes did not simply oppose one another.

For trustworthiness, applying the original prototype or the opposite pattern also increased and decreased trustworthiness as predicted, but only significantly for the latter. These effects were observable on new recordings of the words "hello," but not on other two-syllable words. In every other tested conditions (mean values and dominance filters), perceived trustworthiness decreased. Further analyses revealed, first, that, contrary to dominance, the relation between mean pitch and trustworthiness was nonlinear (Fig. S4*C*), with reduced rather than increased ratings for large mean pitch levels. Second, reverse-correlation analysis on data from the second experiment (*Supporting Information*), suggested that the shape of the trustworthiness prototypes were sensibly different across words (Fig. S4*F*).

Finally, in both tasks, no effects or interactions between listener and stimulus gender were found (all $P > 0.05$), confirming our finding that male and female listeners use similar strategies to process social dominance and trustworthiness in male and female voices.

Discussion

This study demonstrates that social judgments of dominance and trustworthiness from spoken utterances are driven by robust mental prototypes of pitch contours, using a code that is identical across sender and observer gender, and that prosodic mental representations such as these can be uncovered with a technique combining state-of-the-art pitch manipulations and psychophysical reverse correlation.

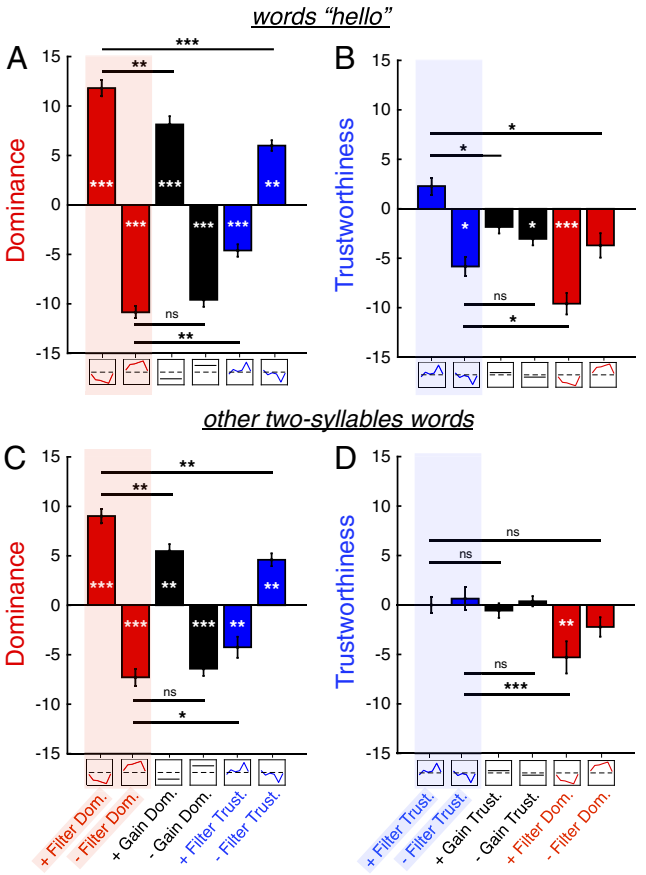

Fig. 3. Applying pitch prototypes as "social makeup" to color novel utterances. Normalized dominance (dom.; *A* and *C*) and trustworthiness (trust; *B* and *D*) judgments (ratings – baseline) obtained by applying the pitch prototypes obtained in the first experiment to new recordings of "bonjour" (*A* and *B*) and other two-syllable words (*C* and *D*). Shaded area highlights the main condition in each task (original prototypes); other conditions (constant gain and other tasks prototype) were tested for control. Planned *t* tests: one-sample *t* tests (vs. 0), white symbols; paired-sample *t* tests (between conditions), black symbols. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$ after Bonferroni correction ($n = 20$ in each task). ns, not significant. Error bars, SEM.

The mental representations found here for dominant prosody, which combine lower mean pitch with a decreasing dynamical pattern, are consistent with previous research showing that people's judgments of dominance can be affected by average pitch and pitch variability (19, 20). Similarly, trustworthy prosodic prototypes, which combine a moderate increase of mean pitch with an upward dynamical pattern, are consistent with findings that high pitch, as well as, e.g., slow articulation rate and smiling voice, increase trusting behaviors toward the speaker (ref. 21; but see ref. 8). Beyond mean pitch, the temporal dynamics of the patterns found here were also consistent with previous associations found between general pitch variations and personality or attitudinal impressions, e.g., falling pitch in assertive utterances (22) or rising pitch in affiliatory infant-directed speech (23). However, the present results show that mental representations for a speaker's dominance or trustworthiness can and should be described in much finer temporal terms than a general rising or falling pitch variation. First, our participants were in striking agreement on shapes sampled at <100 ms (Fig. S3) and the fine details of these shapes, while they generalized to a variety of utterances and (in the

case of dominance) even to other two-syllable words, varied depending on the morphology of the words (Fig. S4). Second, participants gave significantly better evaluations of, e.g., dominance for falling-pitch profiles that were prototype-specific, rather than obtained by inverting the profile of the other construct (Fig. 3).

The fact that both male and female participants relied on the same dynamic pitch prototypes to perceive dominance and trustworthiness in speech is in striking contrast to previous findings of gender effects on vocal dominance judgments (6, 8), and, more generally, the sexual dimorphic features of the human voice (24). Our paradigm, in which pitch variations are generated algorithmically based on otherwise flat-pitch utterances, is able to control for incident variations of male and female prosody that may have obfuscated this processing similarity in previous studies. This finding, which provides behavioral evidence of a unique code for intonation, is consistent with a recent study suggesting that intonation processing is rooted at early processing stages in the auditory cortex (25). Gender symmetry, and more generally independence from a speaker's physical characteristics, seems a very desirable property of a code governing social trait attribution: For instance, judgments of voice attractiveness, which increases via averaging, are also highly similar across gender (26). By focusing on temporal variations in addition to static pitch level, the prosodic code uncovered here appears to be a particularly robust strategy, enabling listeners to discriminate, e.g., dominant from submissive males, even at a similarly low pitch.

While our study shows that both dimensions have distinct prosodic prototypes that are robust within and across participants, the amount to which prototypes inferred on a given word explained the responses on other words differed, with better explanatory power for dominance than trustworthiness. First, it is possible that the trustworthiness prototype, because it appears to be more finely dynamic and tuned to the temporal morphology of the original two-syllable word, is more discriminating of any acoustic–phonetic deviations from this pattern than the smoother dominance kernel. Second, it is also possible that the position of a given exemplar with regard to the prototype is exploited more conservatively in the case of trustworthiness than dominance. In particular, the analysis of response probabilities as a function of the mean pitch change in experiment 2 (Exp. 2) (Fig. S4*C*) shows a more strongly nonlinear relationship in the case of trustworthiness—suggesting that there is such a thing as being "too trustworthy." This pattern is consistent with a recent series of neuroimaging results showing nonlinear amygdala responses to both highly trustworthy and highly untrustworthy faces relative to neutral (27, 28), as well as, behaviorally, more negative face evaluations the more they deviate from a learned central tendency for trustworthiness, but less so for dominance (29).

Given the simple and repetitive nature of the judgment tasks, it appears important to consider whether some degree of participant learning or demand may be involved in the present results. First, one should note that, in Exp. 1, the same intonation pattern was never presented twice. On the contrary, we presented several thousands of different, random intonation patterns across the experiment, in such a way that the experimenters did not a priori favor one shape over another. In Exp. 2, prototypes inferred from Exp. 1 were repeated, but also interleaved with random variations. Therefore, it is unlikely that participants were able to discover, then respond differentially to one particular pitch pattern as the experiment unfolded (see also Fig. S5). While this does not exclude the possibility that participants have set themselves an arbitrary response criteria from the onset of the experiment, this criteria can in no way be guided by conditions decided in advance by the experimenter. Second, because dominance and trustworthiness tasks were conducted on an independent group of participants, the opposite (although nonsymmetric) patterns found for the two constructs cannot be attributed to transfer effects from one task to the other (30). The question remains, however, whether the prototypes evoked in explicit tasks such as the ones described here are consciously accessible to the participants and whether they are similar to those prototypes used in computations in which the corresponding traits are involved, but not directly assessed (see, e.g., ref. 31).

These findings, and the associated technique, bring the power of reverse-correlation methods to the vast domain of speech prosody and thus open avenues of research in communicative behavior and social cognition. First, while these results were derived by using single-word utterances, they initiate a research program to explore how they would scale up to multiword utterances and, more generally, how expressive intonation interacts with aspects of a sentence such as its length, syntax, and semantics. Analyses of infant utterances at the end of the single-word period (32) suggest that prosodic profiles are stretched, rather than repeated, over successive words. Whether such production patterns are reflected in listeners' mental representations can be tested with our technique by using multiword utterances manipulated with single-word filters that are either repeated or scaled to the duration of the excerpts. Another related question concerns how social intonation codes interact with the position of focus words or with conjoint syntactic intonation, both of which are also conveyed with pitch. For instance, English speakers required to maintain focus on certain words may eliminate emotional f0 distinctions at these locations (33). These interactions can be studied with our technique by using reverse correlation on baseline sentences which, contrary to the flat-pitch stimuli used here, already feature prosodic variations or focus markers.

Second, although arguably most important, suprasegmental pitch variations are not the only constitutive elements of expressive prosody, which also affects an utterance's amplitude envelope, speech rate, rhythm, and voice quality (34). By applying not only random pitch changes on each temporal segment, but also loudness, rate, and timbre changes (35), our paradigm can be extended to reveal listeners' mental representations of social prosody along these other auditory characteristics and, more generally, probe contour processing in the human auditory system for other dimensions than pitch, such as loudness and timbre (36). Similarly, while judgements of dominance and trustworthiness may be of prime importance in the context of encounters with strangers, in intragroup interactions with familiar others, e.g., in parent–infant dyads, it may be more important to evaluate states, such as the other's emotions (e.g., being happy, angry, or sad) or attitudes (e.g., being critical, impressed, or ironic). Our method can be applied to all of these categories.

By measuring how any given individual's or population's mental representations may differ from the generic code, data-driven paradigms have been especially important in studying individual or cultural differences in face (13, 16) or lexical processing (37). By providing a similar paradigm to map mental representations in the vast domain of speech prosody, the present technique opens avenues to explore, e.g., dysprosody and social-cognitive deficits in autism spectrum disorder (38), schizophrenia (39), or congenital amusia (40), as well as cultural differences in social and affective prosody (41).

Finally, once derived experimentally with our paradigm, pitch prototypes can be reapplied to novel recordings as social makeup so as to modulate how they are socially processed, while preserving their nonprosodic characteristics such as speaker identity. This process provides a principled and effective way to manipulate personality impressions from arbitrary spoken utterances and could form the foundation of future audio algorithms for social signal processing and human–computer interaction (42).

Materials and Methods

Ethics. All experiments presented in this paper were approved by the "Institut Européen d'Administration des Affaires" (INSEAD) IRB. Participants gave their informed written consent before the experiment, were compensated for participating, and were debriefed and informed about the true purpose of the research immediately after the experiment.

Recordings and Experimental Apparatus. Original stimuli were recorded by a group of native French speakers (see words and speaker characteristics in Table S1), using a DPA 4066 omnidirectional microphone and a RME Fireface 800 soundcard (44.1 kHz) in a double-walled IAC sound-insulated booth. They were mono sound files generated at sampling rate 44.1 kHz in 16-bit resolution by Matlab and were normalized in loudness in the range 75- to 80-dB sound pressure level by using the ITU-R BS.1770 normalization procedure. They were presented diotically through headphones (Beyerdynamic DT 770 PRO; 80 ohms), and sound levels were measured by using a Brüel & Kjær 2238 Mediator sound-level meter placed at a distance of 4 cm from the right earphone.

Voice-Processing Algorithm. We developed an open-source toolbox (CLEESE; available at cream.ircam.fr) to generate the stimuli used in the study. The toolbox, based on the phase-vocoder sound-processing algorithm (43), operates by generating a set of breakpoints (e.g., at every 100 ms in the file) and applying a different audio transformation to every segment. Here, we used the toolbox to set random fundamental frequency values at each of the breakpoints, and linearly interpolate these values within each segment, (see details below for each experiment). Beyond pitch, the toolbox is also able to generate random modifications of spectral envelopes, speed, loudness, and equalization; these additional features were not used in this study.

Validation Experiment. The stimulus used for this experiment was the word "really" ("vraiment" in French), recorded by one male speaker (duration, 426 ms; mean f0, 105 Hz), which can be experienced either as declarative or interrogative depending on its pitch contour. The pitch contour of the utterance was flattened, then divided into six segments of 71 ms; we manipulated the pitch over the seven time points on Gaussian distributions of SD = 70 cents, clipped at $\pm$ 2.2 SD. Pairs of these randomly manipulated voices were presented to five observers who were asked, on each trial, to judge which of the two versions appeared most interrogative. Each observer was presented with 700 trials. Prosodic prototypes were computed as described in Fig. 1 (see also Exp. 1 and below).

Exp. 1.

Participants. A total of 21 participants (female: 10; M = 21) were in the dominance task, and 23 participants (female: 11; M = 22) were in the trustworthiness task. All were native French speakers with normal hearing.

Procedure. Participants listened to a pair of two randomly modulated voices and were asked which of the two versions was most dominant (in one participant group) or most trustworthy (in the other group; examples of trials are in Audio File S1). The interstimulus interval was 500 ms, and the intertrial interval was 1 s. In each session, participants were presented with a total of ~700 trials. Participants took part in two sessions, which took place on different days: one session with male voices and the other with female voices (counterbalanced between participants).

Stimuli. We created the stimuli by artificially manipulating the pitch contour of one male and one female recording. First, the pitch contour of both recordings was artificially flattened to constant pitch, by using the processing shown in Fig. S2. Then, we added/subtracted a constant pitch gain ($\pm$20 cents, equating to $\pm$1 fifth of a semitone) to create the "high-pitch" or "low-pitch" interval in each 2I-2AFC trial. Finally, we added Gaussian "pitch noise" to the contour by sampling pitch values at six successive time points, using a normal distribution (SD = 70 cents; clipped at $\pm$ 2.2 SD). These values were linearly interpolated between time points and fed to a pitch-shifting toolbox developed for this purpose (see above).

Statistical analyses. All tests were two-tailed and used the 0.05 significance threshold. Huynh–Feldt corrections for degrees of freedom and Holm–Bonferroni corrections for multiple measures were used where appropriate.

Mean pitch difference analysis. To assess how mean pitch drove judgements of dominance/trustworthiness, we computed the mean pitch difference in each pair of voices. For each listener, we divided these measures into 15 equal sets and computed the response probability for dominant/trustworthy in each task. We fitted psychometric functions to the data of each listener using logistic functions: $f(x) = (a0 + b0./(1 + \exp(-(x - x0)^{*} s0))$ and the nonlinear least-squares method of the Matlab "fit" function. Lower and upper bounds for the regression parameters [$a0$, $b0$, $s0$] were [-20, -30, -10] and [20, 30, 10], respectively; the initial condition was [1, 0, 20]. The different conditions (listener and stimuli gender) were compared by using two 2 $\times$ 2 [SubGender $\times$ StimGender] ANOVAs (one per task) on the slope values ($s0$) from these regressions. For the dominance task, we found no effect of listener gender [$F(1, 19) = 3.53$, $P > 0.05$, $\eta_p^2 = 0.16$], a significant effect of stimuli gender [$F(1, 19) = 8.61$, $P = 0.009$, $\eta_p^2 = 0.31$], and no interaction between listener and stimuli gender [$F(1, 19) = 1.16$, $P > 0.05$, $\eta_p^2 = 0.06$]. The slopes were slightly but significantly steeper for the male voice than the female voice. For the trustworthiness task, we found no effect of listener gender [$F(1, 21) = 1.02$, $P > 0.05$, $\eta_p^2 = 0.05$], no effect of stimuli gender [$F(1, 21) = 0.44$, $P > 0.05$, $\eta_p^2 = 0.02$], and no interaction between listener and stimuli gender [$F(1, 21) = 4.10$, $P > 0.05$, $\eta_p^2 = 0.16$].

Reverse-correlation analysis. A first-order temporal kernel (18) was computed for each subject in each session, as the mean pitch contour of the voices classified as dominant [respectively (resp.) trustworthy] minus the mean pitch contour of the voices classified as nondominant (resp. nontrustworthy). Kernels were then normalized in each condition by dividing them by the sum of their absolute values (44) and then averaged in each task [as there is no consensus yet in how to "aggregate" individual kernels, i.e., whether to normalize individual kernels or not (45), we also replicated the analysis using raw kernels and verified that conclusions remained unchanged]. A mixed 6 $\times$ 2 $\times$ 2 (Segment $\times$ StimGender $\times$ SubjGender) repeated-measures ANOVA (rmANOVA) was conducted on the temporal kernels obtained in each task separately.

Dominance task. The effect of segment was significant [$F(5, 95) = 17.44$, $P < 0.001$, $\eta_p^2 = 0.48$, $\tilde{\varepsilon} = 0.47$]. A significant Segment $\times$ StimGender interaction was also obtained [$F(5, 95) = 5.09$, $P < 0.001$, $\eta_p^2 = 0.21$, $\tilde{\varepsilon} = 0.86$]. Post hoc t tests revealed a significant difference between the filters for the male and the female voice on the fourth segment, which plausibly resulted from close but distinct intrasyllabic loudness contours between the two voices. Neither significant effect of participants' gender nor any interaction with other variables was found ($P > 0.05$).

Trustworthiness task. The effect of segment was significant [$F(5, 105) = 9.27$, $P < 0.001$, $\eta_p^2 = 0.31$, $\tilde{\varepsilon} = 0.53$]. There was no other significant main effect or interaction with other variables ($P > 0.05$).

Exp. 2.

Participants. A total of 21 participants (female: 9; M = 22) were in the dominance task, and 19 participants (female: 10; M = 21) were in the trustworthiness task. Participants were native French speakers with normal hearing.

Procedure. Each group gave his or her ratings on a Likert scale ranging from extremely nondominant/nontrustworthy to extremely dominant/trustworthy. Subjects were presented with 420 stimuli [20 words $\times$ 7 conditions/filters $\times$ 3 repetitions; a few (<3) randomly missed trials] in three 20-min blocks. All of the words and conditions were randomly mixed. Participants received false performance feedback (as a random score ranging between 70 and 90/100) at the end of each block to enforce concentration.

Stimuli. We applied seven different filters to the pitch contours of a set of 20 recordings (all with flattened pitch; see Fig. S2 for details). These filters were: (*i*) four dynamic filters derived from Exp. 1 (both dominance and trustworthiness kernels and their opposite patterns multiplied by 1,050 cents); and (*ii*) three other static filters, used as control: a baseline condition with no change and two constant-gain filters calculated as the mean value of the kernels from Exp. 1. On these three constant kernels, we applied the same amount of random pitch fluctuations as the kernels of Exp. 1, to avoid stimuli with unrealistically flat intonations; this also allowed us to replicate reverse-correlation analysis on this new set of data (see below). The value of 1,050 cents (=15 SD used in the task) was chosen to generate pitch changes with peaks ~250 cents.

Comparison across conditions analyses. For each subject and each stimulus, we computed the mean rating in each condition and divided this value by the mean rating collected in the baseline (no change + noise) condition. This procedure discarded effects of the speaker and the word. Two mixed 8 $\times$ 2 $\times$ 2 $\times$ 2 [Condition $\times$ SubGender $\times$ StimGender $\times$ TypeStim] ANOVAs were conducted on the normalized ratings in the dominance and the trustworthiness task, separately. In the dominance task, we found a significant effect of the condition [$F(5, 95) = 54.6$, $P < 0.001$, $\eta_p^2 = 0.74$, $\tilde{\varepsilon} = 0.24$] and a Condition $\times$ TypeStim interaction [$F(5, 95) = 6.6$, $P < 0.001$, $\eta_p^2 = 0.26$, $\tilde{\varepsilon} = 0.56$]. In the trustworthiness task, we found a significant effect

of the condition [$F(5, 85) = 8.8$, $P = 0.002$, $\eta_p^2 = 0.34$, $\tilde{\varepsilon} = 0.35$], an effect of the type of stimuli used TypeStim [$F(1, 17) = 7.1$, $P = 0.02$, $\eta_p^2 = 0.29$], and a Condition*TypeStim interaction [$F(5, 85) = 8.6$, $P < 0.001$, $\eta_p^2 = 0.34$, $\tilde{\varepsilon} = 0.81$].

Reverse-correlation analysis. As a supplementary analysis, we also subjected data from Exp. 2 to reverse correlation. First-order temporal kernels were computed for each subject and each stimulus by using the trials of the three conditions where noise was added on the contour (i.e., baseline and constant gains). The trials were coded as dominant/nondominant or trustworthy/nontrustworthy if they were higher/lower than the mean of the ratings produced by the subject for this particular stimulus. Statistical analyses and averaged kernels for each group are presented in Fig. S4.

ACKNOWLEDGMENTS. The work was supported by European Research Council Starting Grant ERC CREAM Grant 335536 (to J.-J.A.); and by Institute of Language, Communication and the Brain Grant ANR-16-CONV-0002, Brain and Language Research Institute Grant ANR-11-LABX-0036, and A*MIDEX Grant ANR-11-IDEX-0001-02 (to P.B.). Funding was also provided by Grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL* from the Agence Nationale de la Recherche. Data were collected at the Centre Multidisciplinaire des Sciences Comportementales Sorbonne Universités-INSEAD.

1. Ambady N, Rosenthal R (1992) Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychol Bull* 111:256–274.
2. Wojciszke B, Bazinska R, Jaworski M (1998) On the dominance of moral categories in impression formation. *Pers Soc Psychol Bull* 24:1251–1263.
3. Fiske ST, Cuddy AJ, Glick P (2007) Universal dimensions of social cognition: Warmth and competence. *Trends Cogn Sci* 11:77–83.
4. Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. *Proc Natl Acad Sci USA* 105:11087–11092.
5. Jack RE, Schyns PG (2017) Toward a social psychophysics of face communication. *Annu Rev Psychol* 68:269–297.
6. McAleer P, Todorov A, Belin P (2014) How do you say "hello"? Personality impressions from brief novel voices. *PLoS One* 9:e90779.
7. Feinberg DR, Jones BC, Little AC, Burt DM, Perrett DI (2005) Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Anim Behav* 69:561–568.
8. Tsantani MS, Belin P, Paterson HM, McAleer P (2016) Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception* 45:946–963.
9. Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 70:614–636.
10. Barr DJ (2003) Paralinguistic correlates of conceptual structure. *Psychon Bull Rev* 10:462–467.
11. Mitchell RL, Ross ED (2013) Attitudinal prosody: What we know and directions for future study. *Neurosci Biobehav Rev* 37:471–479.
12. Ohala JJ (1984) An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica* 41:1–16.
13. Adolphs R, et al. (2005) A mechanism for impaired fear recognition after amygdala damage. *Nature* 433:68–72.
14. Adolphs R, Nummenmaa L, Todorov A, Haxby JV (2015) Data-driven approaches in the investigation of social perception. *Phil Trans R Soc B* 371:20150367.
15. Dotsch R, Todorov A (2012) Reverse correlating social face perception. *Soc Psychol Pers Sci* 3:562–571.
16. Jack RE, Garrod OG, Yu H, Caldara R, Schyns PG (2012) Facial expressions of emotion are not culturally universal. *Proc Natl Acad Sci USA* 109:7241–7244.
17. Murray RF (2011) Classification images: A review. *J Vis* 11:1–25.
18. Ahumada A, Jr, Lovell J (1971) Stimulus features in signal detection. *J Acoust Soc Am* 49:1751–1756.
19. Hodges-Simeon CR, Gaulin SJ, Puts DA (2010) Different vocal parameters predict perceptions of dominance and attractiveness. *Hum Nat* 21:406–427.
20. Watkins CD, Pisanski K (2017) Vocal indicators of dominance. *Encyclopedia of Evolutionary Psychological Science*, eds Shackelford TK, Weekes-Shackelford VA (Springer, Cham, Switzerland), pp 1–6.
21. Torre I, White L, Goslin J (2016) Behavioural mediation of prosodic cues to implicit judgements of trustworthiness. *Proceedings of Speech Prosody 2016* (International Speech Communication Association, Baixas, France), pp 816–820.
22. Fernald A (1989) Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Dev* 60:1497–1510.
23. Grieser DL, Kuhl PK (1988) Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Dev Psychol* 24:14–20.
24. Titze IR (2000) *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA).
25. Tang C, Hamilton L, Chang E (2017) Intonational speech prosody encoding in the human auditory cortex. *Science* 357:797–801.
26. Bruckert L, et al. (2010) Vocal attractiveness increases by averaging. *Curr Biol* 20:116–120.
27. Todorov A, Baron SG, Oosterhof NN (2008) Evaluating face trustworthiness: A model based approach. *Soc Cogn Affect Neurosci* 3:119–127.
28. Freeman JB, Stolier RM, Ingbretsen ZA, Hehman EA (2014) Amygdala responsivity to high-level social information from unseen faces. *J Neurosci* 34:10573–10581.
29. Dotsch R, Hassin RR, Todorov A (2016) Statistical learning shapes face evaluation. *Nat Hum Behav* 1:0001.
30. Poulton EC (1982) Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychol Bull* 91:673–690.
31. Young AI, Ratner KG, Fazio RH (2014) Political attitudes bias the mental representation of a presidential candidate's face. *Psychol Sci* 25:503–510.
32. Branigan G (1979) Some reasons why successive single word utterances are not. *J Child Lang* 6:411–421.
33. Pell MD (2001) Influence of emotion and focus location on prosody in matched statements and questions. *The J Acoust Soc Am* 109:1668–1680.
34. Grandjean D, Bänziger T, Scherer KR (2006) Intonation as an interface between language and affect. *Prog Brain Res* 156:235–247.
35. Ponsot E, Arias P, Aucouturier J (2018) Uncovering mental representations of smiled speech using reverse correlation. *J Acoust Soc Am* 143:EL19–EL24.
36. McDermott JH, Lehr AJ, Oxenham AJ (2008) Is relative pitch specific to pitch? *Psychol Sci* 19:1263–1271.
37. Varnet L, Wang T, Peter C, Meunier F, Hoen M (2015) How musical expertise shapes speech perception: Evidence from auditory classification images. *Sci Rep* 5:14489.
38. Jiang J, Liu X, Wan X, Jiang C (2015) Perception of melodic contour and intonation in autism spectrum disorder: Evidence from Mandarin speakers. *J Autism Dev Disord* 45:2067–2075.
39. Pinheiro AP, et al. (2013) Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: An electrophysiological investigation. *Psychol Med* 43:603–618.
40. Liu F, Patel AD, Fourcin A, Stewart L (2010) Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain* 133:1682–1693.
41. Sauter DA, Eisner F, Ekman P, Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc Natl Acad Sci USA* 107:2408–2412.
42. Vinciarelli A, Pantic M, Bourlard H (2009) Social signal processing: Survey of an emerging domain. *Image Vis Comput* 27:1743–1759.
43. Liuni M, Roebel A (2013) Phase vocoder and beyond. *Musica Tecnol* 7:73–120.
44. Oberfeld D, Plank T (2011) The temporal weighting of loudness: Effects of the level profile. *Atten Percept Psychophys* 73:189–208.
45. Neri P, Levi D (2008) Evidence for joint encoding of motion and disparity in human visual perception. *J Neurophysiol* 100:3117–3133.