ORIGINAL ARTICLE

Effects of Secondhand Information on Impression Formation in Spoken Communication

Miho KITAMURA*,** and Katumi WATANABE***,****

* Waseda Institute for Advanced Study, 1-21-1 Nishiwasda, Shinjuku-ku, Tokyo 169-0051, Japan
** IdeaLab. Inc., 3-29-1 Ebisu, Shibuya-ku, Tokyo 150-0013, Japan
*** Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
**** University of New South Wales, NSW 2052, Australia

Abstract: This study investigated the role of secondhand information for impression formation in spoken communication, using a voice transfer system. In Experiment 1, listeners were presented with spoken sentences that represented someone’s behavior in either a positive or negative prosody, which is a clue for forming impressions. Listeners rated each social impression of a person who took on a specific behavior. The results showed that the formed impressions were weaker when the affective characteristics of voices were inconsistent with those of behavioral descriptions. Experiment 2 tested the effects of sentence meanings on forming impressions of speakers’ voices using the same sound files. Results revealed that social impressions of speakers’ voices were also modulated by sentence meaning, even though the sentence was not related to the speaker. This study underscores the role of secondhand information in social impression and advances the understanding of the interaction between prosody and meanings in spoken communication.

Keywords: Impression formation, Voice, Communication, Prosody

1. INTRODUCTION

Forming impressions is the process of integrating pieces of information about people and forming a global impression of them. We are quick to establish lasting impressions of others based on their current behavior and appearance [1-4]. We can also form impressions of others even when we do not see someone in person. We often hear from others about what someone is like. For example, a colleague may tell you what the new manager is like. Based on such a verbal description, we often make a first impression and decide what our future behavioral agenda will be. This study focuses on the process of creating impressions based on what the listener is told about others.

Ames et al. [5] suggested that there are three main ways of forming an impression of others: (1) secondhand information (being told about someone); (2) direct behavioral experience (interacting with someone); and (3) appearance (seeing someone in person). Our desire to understand social things often exceeds the limits of what we can absorb through direct experience. Since over 60% of conversation is related to social topics [6], social impressions are likely to often be made via secondhand information [5]. Although much research has been focused on direct impression formation, in daily life, social impressions are often made through secondhand information of a target person who is not present.

Secondhand impressions differ in many ways from direct ones (e.g., views developed based on appearance or direct behavioral experiences), because secondhand information is conveyed by a third person [7, 8]. One of the differences is that secondhand impressions tend to be less accurate than direct ones because the latter are usually based on more reliable details [7]. The most interesting difference between secondhand information and other ways of forming impressions is that the former includes the speaker’s attitude toward that person.

For spoken communication, speakers change not only what they say but also how they speak to represent their emotional attitude. For example, when a speaker says “That guy is nice” in a harsh tone of voice, the listener is likely to infer that the speaker does not think so and wishes to convey something else. In other words, in this situation, prosody does not need to be consistent with the sentence meaning. Impressions from secondhand information can be induced by the interaction of what the speaker said and how he/she said it. Although there are many studies about the effects of voice on forming direct impressions (e.g., voice pitch is related to trustworthiness and attractiveness [9-11] or people can also construct first impressions of speakers from pieces of a voice [12]), it is unclear how people make social impressions from secondhand information, as well as how they integrate what the speaker says and how he/she speaks about a third person.

To understand this process, the present study examined
whether the way in which one speaks modulates impression formation based on secondhand information; we did so by adjusting prosody through an advanced voice transformation system [13].

There is little research on how the way in which one speaks interacts with speech meaning during impression formation (including direct and in-direct formation). However, it is known that listeners integrate speech meanings and voice prosody to identify a speaker’s emotion [14-23]. For example, prosodic information can help resolve the ambiguity of the meanings of spoken words [23], and prosody speeds up naming responses to spoken words [22]. These suggest that prosody and sentence meanings interact in emotion inference processes.

Taken together, it is highly likely that prosody also plays a vital role in forming impressions of the third person during spoken conversations.

Most previous emotion recognition studies have used short, simple words to examine the interactions between sentence meanings and prosody and have demonstrated powerful interactions between them. One of the reasons why they employed short words is that it is difficult to produce long sentences while maintaining emotional validity. Several studies have tried to create them [14, 22, 23]; for example, some researchers generated their own datasets of sound stimuli, maintaining validation of the emotionality of their acoustic stimuli. They usually recruited well-trained actors to record emotional sentences and confirmed the validation of the emotional properties of these sounds’ sentence meanings and prosody by scrutinizing the judgments of many different participants. Although this procedure for creating stimuli is widely used, the use of such datasets has several limitations [24]. The most important problem is that actor-recorded stimuli do not allow for fine control over the intensity with which feelings are expressed; actors may be more expressive when it comes to happiness than sadness, and there could be expressional variations between actors. To overcome this problem, researchers often employ morphing techniques between two different emotional vocal expressions [25]. However, morphing techniques affect not only emotional cues, but also non-emotional signals that researchers do not want to change. For example, it is technically impossible to only modify the pitch—but not the loudness—between two emotional voices.

To resolve these issues, Aucouturier et al. [13, 24] recently developed a real-time voice modulation system called Da Amazing Voice Inflection Device, or DAVID. This system allows us to create an emotional voice for any kind of speech stimuli. DAVID is a collection of “audio effects” that can be combined in various changes of acoustic features to transform voice toward a specific emotion [13, 24]. Each audio effect corresponds to a frequently identified correlation of an emotional voice in a real emotional situation [24]. For instance, fear is often associated with fluctuations in pitch in natural scene.

To create natural emotional voices, DAVID contains four components that modulate voice pitch, inflection, vibration, and frequency. By combining these elements, DAVID can generate several emotional voices, such as happy, sad, and afraid. For instance, the happy manipulation alters the pitch of the speaker’s voice by using upshifting and inflection to make it seem more positive [24]. The validity of this manipulation has been confirmed in several languages [24]. The amount of physical changes of each acoustic parameter (pitch shift, vibrato, inflection, and filter) corresponding to each emotion has been shown in Aucouturier et al.’s studies in more detail [13, 24]. In the present study, the voices were changed toward a specific affective direction (positive or negative) using the DAVID system, and it was investigated how prosodic affective modulation influences social impressions of a third person based on secondhand information received during spoken communication.

Previous research examining interactions between prosody and sentence meanings in relation to emotion recognition used direct emotional sentences that describe or evoke feelings, such as “This song makes me cry” [14, 22, 23]. Social impressions are usually developed based on a diverse array of present and perceptible behaviors (except for physical appearance); people use these impressions to expect future behaviors from others. Since present study focused on the processes of impression formation, instead of emotional sentences, we used person-behavior descriptive sentences that imply certain personalities (i.e., facilitating impression formation) [26]. These sentences only describe a person’s behavior (not feelings). For example, one of the positive sentences was, “He invited the newcomers to his house.” Using these scripts, we were able to directly analyze the processes of impression formation via secondhand details (versus the processes of emotion recognition).

Classical empirical studies of social impressions have examined impression formation from secondhand information, such as verbal descriptions [27]. Participants were given a list of traits that contained adjectives such as warm, cold, intelligent, and cautious, and they evaluated the impressions of that person [28-31]. Such research has provided the fundamental proposition and general principles of forming impressions (for a review, see [32]).
The underlying theory is that observers assume unity in others’ personalities and view them as coherent entities. Impression formation is the process of integrating information about a person to create a coherent opinion. Based on this notion, we expected that if prosody and behavioral descriptions were to have inconsistent qualities, listeners would integrate or average each impression of prosody and the behavioral description, even though they are derived from different sources (Experiment 1).

If prosody and sentence meanings both play an important role in establishing social impressions, then perceptions of speakers’ voices would also be influenced by sentence meanings. Regarding this notion, several studies have reported that it is difficult for listeners to ignore irrelevant sentence meanings when asked to focus on prosody when carrying out an emotion categorization task [18, 19]. These previous studies have shown that spoken words and prosody interact with each other because they are derived from the same person. However, in secondhand impression formation, the sentence meaning is not related to the speaker because it is about a third person. In other words, it is not necessary for the listeners to integrate voice prosody and speech meanings to form impressions about the speaker. In Experiment 2, using the same sound set as in Experiment 1, we studied whether impressions of a speaker’s voice are modulated by sentence meaning that is unrelated to the speaker.

2. EXPERIMENT 1: The influence of a speaker’s affective prosody on secondhand impression formation

2.1 Method

2.1.1 Participants

Twenty-six participants (13 females and 13 males) were recruited for Experiment 1. The mean age was 21.3 years, with a standard deviation (SD) of 1.63 (range: 18 to 25 years old). All participants were native Japanese speakers and reported normal hearing. They were unaware of the purpose of the experiment. Our experiment was approved by the ethics committee at Waseda University. All subjects gave informed consent and were paid for their participation.

2.1.2 Stimuli

We prepared a total of 54 sentences that described specific behaviors facilitating impression formation (Supplementary Information). For example, one of the positive descriptions was, “He invited the newcomers to his house.” In this case, most people would form positive impressions of this person. A total of 89 person-descriptive sentences were translated [26] into Japanese and pre-tested (N = 39) for valence (1 = very negative and 5 = very positive) and arousal (1 = not arousing and 5 = very rousing). We then calculated the average rating scores for each sentence and chose 18 sentences that got close scores for positive (max: 5), neutral (3), and negative (min: 1) characteristics. The total number of sentences considered was 54. We conducted a one-way repeated measures analysis of variance (ANOVA) among the mean arousal ratings of positive, neutral, and negative sentences to ensure that there were no significant arousal differences (p > 0.1). A female native Japanese speaker recorded these sentences using an omnidirectional headset microphone (DPA d:Fine 4066) connected to an Apple Macintosh (Mac mini) computer. Digitization of the recordings was done at a 44,100 Hz sampling rate and 16-bit resolution. We also pre-tested the original voice stimuli to confirm that the voice stimuli describing various affective behaviors would convey different impressions to listeners. Another group of 16 participants (9 females, mean age = 20.19 years [SD = 3.02]) listened to these original voice stimuli and rated their impressions of the speaker, who took on the behaviors described by the voice stimuli. The baseline affective characteristics of the voice stimuli are shown in Table 1.

Next, we processed the recorded digital audio files using the DAVID (version 1.1) software to give the voice happy (i.e., positive) and sad (i.e., negative) tones for the 54 behavioral sentences. As we mentioned in Section 1, the DAVID software can manipulate voice pitch, inflection, vibration, and frequency. To induce positive and negative voice, we set the default transfer mode of happy and sad voice in DAVID and recorded our voice stimuli in this setting. The amount of change in each acoustic feature (pitch shift, vibrato, inflection, and filter) was shown in [13, 24]. Please see [13, 24] for the details. Previous research showed that this emotional transformation has yielded above-chance recognition and perception of naturalness [24].

The total recorded voices were 108 (54 sentences: [18 positive, 18 neutral, 18 negative sentences: see Supplementary Information] × 2 voices [happy and sad]). For the experimental trials, the 108 voice stimuli were divided into two sets of 54. Since we were interested in first impressions, each sentence was presented only once.

### Table 1: Results of the pre-test that examined affective impressions of original voice stimuli (N=16)

| Category        | Average scores | SD |
|-----------------|----------------|----|
| Positive descriptions | 4.05 | 0.97 |
| Neutral descriptions | 2.77 | 1.21 |
| Negative descriptions | 1.95 | 1.07 |
Therefore, two experimental versions of the stimuli were created by counterbalancing voice prosody. For example, nine positive descriptions were presented in the happy voice in one version and in the sad voice in the second version, while the other nine descriptions were presented in the sad voice in one version and in the happy voice in the second version. Participants were randomly assigned to one of the two experimental versions. Therefore, for each participant, 27 sentences (positive, neutral, and negative) were presented in a happy voice, and the other 27 sentences were presented in a sad voice. There were non-significant differences in the valence scores between the two versions ($t(53) = 0.05, p = .956, n.s.$).

2.1.3 Procedure and apparatus

The experiment was carried out in a dimly lit room. Auditory stimuli were presented to participants through open headphones (SENNHEISER HD 380 PRO). We controlled the experiment and collected the data using MATLAB® (The MathWorks) with the Psychophysics toolbox extension [33,34] running on an Apple Macintosh (Mac mini) computer.

Participants were informed that they would listen to sound files describing the behaviors of different individuals and that they would be asked to report their impressions of them. As mentioned in Section 2.1.2, to avoid repeating sentences, two versions of the stimulus were prepared by counterbalancing the prosody assigned to each set (happy or sad). For example, we presented each sound with a happy tone in the first version and with a sad tone in the second. Participants were randomly assigned to one of the two experimental versions. For each participant, we presented 27 sentences in a happy voice, and 27 in a sad voice. Thus, the total number of trials was 54 per participant. We randomized the order of trials for each participant.

After the experimenter provided instructions, the participants viewed an impression rating scale ranging from 1 (a negative person) to 5 (a positive person) on a computer display in front of them. After 500 ms, they heard the voice stimulus through their headphones. Then, the participants rated their impressions of the individual who expressed the behaviors that the voice stimulus was describing on the scale. After assigning rating, participants heard the next sound.

2.1.4 Statistical analyses

We averaged the rating scores across the sentences of each behavioral description category for each participant. For the data, we conducted a traditional two-way repeated measures ANOVA with the behavioral description category (positive, neutral, and negative) and prosody of voice (happy and sad) as within-participant factors.

2.2 Results and discussion

Figure 1 shows the mean rating scores as a function of behavioral descriptions spoken with a negative or positive prosody. A two-way repeated measures ANOVA with the behavioral description category (positive, neutral, and negative) and prosody of voice (happy and sad) as within-participant factors revealed that impressions of positive behaviors were generally higher than those of other behaviors ($F(2, 50) = 94.26, p<.01; \text{partial } \eta^2 = .79$). Interestingly, the impression ratings for happy prosody were higher than those of sad prosody ($F(1, 25) = 4.79, p<.05; \text{partial } \eta^2 = .16$), and the interaction was significant ($F(2, 50) = 3.99, p<.05, \text{partial } \eta^2 = .13$). Post-hoc pairwise comparisons with the Bonferroni correction demonstrated that for the negative and positive behaviors spoken with happy prosody, more positive impressions were revealed than behaviors spoken with sad prosody ($p < .05$). Meanwhile, for neutral behaviors, there were no significant differences between happy and sad prosody.

The neutral descriptions did not have specific emotionality in and of themselves. Regarding facial recognition, several studies have revealed that emotionally neutral stimuli are affected by their emotional context [35]. Based on these findings, it is highly likely that...
impressions of neutral behaviors would be more affected by voice. However, our results showed the opposite, suggesting that voice affectivity interacts with the affective meaning of speech.

It was unclear whether a happy voice improved speech impressions or a sad voice worsened them. In our preliminary test (N = 16; these were different participants from those of Experiment 1), we had the baselines of the affective impressions of these behaviors spoken by the original voice (i.e., the non-filtered voice). In Figure 1, three horizontal dotted lines represent the mean baseline rating scores of positive, neutral, and negative behavioral descriptions, respectively. Comparing these baseline scores for the behavioral description category (negative, positive, and neutral) of the original voices with those of the affective voices, we conducted a single sample t-test with baseline scores, or original sound stimuli for each behavior category. It was found that happy voices improved impressions of negative behaviors (t(25) = 1.75, p < .05, d = 0.34, one-tailed). Meanwhile, a sad voice worsened impressions of positive behaviors (t(25) = -3.66, p < .001, d = 0.72, one-tailed).

In Experiment 1, we found that the process of impression formation was modulated by voice prosody. This outcome implies that while forming social impressions in spoken communication, prosody and sentence meanings interact. These results are consistent with previous emotion recognition research that showed that semantic emotion recognition was interfered by voice prosody [16-22]. Moreover, we found that happy voices improved impressions of negative behaviors, while sad voices worsened impressions of positive behaviors. Classical empirical research on social impressions indicates that impression formation involves integrating information about a person to form a coherent impression [32]. In line with this notion, our results illustrated that the listeners averaged prosody and behavioral descriptions to develop coherent impressions, even though they were derived from different sources.

In our second experiment, we were interested in testing the opposite scenario. In other words, whether impressions of the speaker’s voice would be modulated by the sentence meanings describing the behavior of others, that is, content, is not related to the speaker. Perceiving affective properties in a person’s voice and establishing an impression are fundamental aspects of human communication. The voice provides considerable personal information about speakers, such as identity and emotional state [36], and we can quickly form affective impressions of a speaker based on his/her voice [12]. However, it is not enough to examine whether impressions of a speaker’s voice are influenced by what he/she said. Especially for the secondhand impression formation, it is unclear whether voice impressions would be modulated by sentence meanings, because speech content is not directly related to the speaker; in other words, speakers are describing the behavior of others. The second experiment, using the same sound set as in Experiment 1, examined how impressions of the speaker’s voice were modulated by the contents of their speech describing the behavior of others, which are not related to the speaker.

3. EXPERIMENT 2: Effects of behavior descriptions on voice impressions

3.1 Method

3.1.1 Participants

Twenty-six participants (11 females and 15 males) took part in Experiment 2. The mean age was 20.82 years, with an SD of 1.56 (age range: 18 to 25 years old). All participants were native Japanese speakers and reported normal hearing. They were unaware of the purpose of the experiment. Our experiment was approved by the ethics committee at Waseda University. All subjects gave informed consent and were paid for their participation.

3.1.2 Stimuli, apparatus, and procedure

The stimuli, apparatus, and procedure were the same as in Experiment 1, except for the instructions conveyed to the participants. We asked them to rate the affective impressions of the voices they heard. We also asked them to ignore the emotional meanings of the sentences to exclude over-focusing on sentence meanings.

3.1.3 Statistical analyses

We averaged the rating scores across the sentences of each behavioral description category for each participant. For the data, it was conducted a traditional two-way repeated measure ANOVA with prosody of voice (happy and sad) and behavioral description category (positive, neutral, and negative) as within-participant factors.

3.2 Results and discussion

Figure 2 presents the rating scores as a function of prosody in which behavioral descriptions were spoken. A two-way repeated measures ANOVA with prosody of voice (happy and sad) and behavioral description category (positive, neutral, and negative) as within-participant factors showed that happy prosody revealed more positive impressions than sad prosody, regardless of the emotionality of the behavioral descriptions (F(1, 25) = 100.79, p < .01,
Interestingly, the voices that stated positive descriptions were perceived as more positive than the voices that mentioned negative descriptions, irrespective of their voice emotionality $\eta^2 = 0.26$. Contrary to the findings of Experiment 1, the interaction was not significant $F(2, 50) = 0.83$, n.s.). Post-hoc pairwise comparisons with the Bonferroni correction demonstrated that more negative impressions were revealed for negative descriptions than the positive and neutral descriptions, notwithstanding the initial voice emotionality $p < .05$. The results suggest that affective descriptions modulate impressions of a speaker’s voice; in other words, affective contents of spoken sentences, unrelated to speaker’s personalities, changed the listeners’ impressions of the speaker’s voice.

4. GENERAL DISCUSSION

The present study examined how prosody affects impression formation based on secondhand information. The results demonstrated that prosody modulated participants’ views toward suppressing the original personality impressions. In addition, it was found that impressions of the speaker’s voice were influenced by the affective properties of the sentence meaning that was not directly related to the speaker. These results indicate that when listeners form impressions of speakers and others based on auditory information, listeners must integrate information about what the speaker says and how he/she says it.

The results of Experiment 1 revealed that negative behaviors spoken about in a happy voice caused more positive impressions, while positive behaviors spoken about in a sad voice produced more negative impressions than the original ones. These results imply that incongruent prosody suppresses original impressions of others based on their behavioral descriptions. Impression formation is the process of incorporating details about a person to develop a coherent impression [32]. In line with this thinking, we found that when the affective characteristics of voices and sentence meanings were inconsistent with each other, listeners averaged or integrated the traits to form coherent impressions. Previous empirical studies on social impressions have focused on the inference process, averaging different qualities belonging to one target person. For example, “He is painstaking and well-spoken.” In the current study, we explored the formation of impressions based on the speaker’s voice and the behavior of another person and found that listeners were able to integrate affective properties from different sources. These results imply that this integrative approach is a fundamental process of inferring personalities when social information is derived from different sources.

The present study did not find any effects of voice on neutral behavior descriptions; this suggests that voice prosody did not bias the impressions derived from neutral behaviors of the third person. Previous emotion recognition research has shown that awareness of meaning in ambiguous spoken words is biased toward the emotional tone of voice of speakers [20]. In addition, a recent empirical study on direct impressions reported that moderately attractive faces tend to be more affected by language information (i.e., personal profile of that person) than lesser or highly attractive faces [37]. Given that these previous studies investigated the integration process of multiple information from the same person, and not different people, there is a possibility that the present results are due to secondhand information.

In terms of emotion recognition, previous studies found that sentences representing consistent emotions in voice prosody of speakers were rated highest by listeners [14, 38], indicating that listeners added the emotional value of voices to that of sentences. However, the present study did not observe this kind of additive process. When the affective meaning of the voice was consistent with the sentence meaning, the tone of voice did not underscore the affective meanings of the behavioral descriptions. One possible reason for our results would...
be related to quality of emotional information. In the present study, we used behavioral descriptions and prosody. Although participants rated both stimuli using the same affective scale (from positive to negative), there is a possibility that emotional quality would not completely match each other. Our study is the first of its kind to examine the processes of forming impressions by combining voice prosody and behavioral descriptions. Inferring personalities from diverse sources would differ from inferring feelings from varied sources, even though both judgments are based on affective information. In future research, we should investigate the discrepancies between them and specify the functional roles of these systems.

In Experiment 2, it was found that impressions of a speaker’s voice were modulated by sentence meaning, signifying that impression formation of a speaker’s voice would be affected by the contents of the speech, even though these contents were unrelated to the speaker. Many studies on affective information processing have found that comprehension of the affective meaning of sensory inputs (e.g., the face, voice, or posture) is influenced by the situational context, derived from different information sources. For instance, recognition of facial expressions can be influenced by circumstances, such as descriptions of a social situation [39], voice, body posture, and visual scenes [40–42]. In terms of identifying emotions in voices, recognition accuracy for vocal emotions is generally lower (about 55–65%) than that for facial expressions (about 75%; for a review, see [43]). Considering these findings, inferring a person’s affective traits based on his/her voice would be easily influenced by the context; that is, it would be easy to combine the speaker’s voice and sentence meanings and integrate them, even though the speaker has spoken about the behavior of another person.

There are limitations to the current study. First, in the present study, we used happy and sad voices created by DAVID to change how the speaker talks. Although DAVID enables stable control of voice prosody, it cannot support wide variations in how to speak (such as how to convey liking, fear, and anger). Future research should examine the interactions between sentence meanings and prosody that reflect various speakers’ intentions. Second, we only recorded voice stimuli from one Japanese female speaker. This means that the variance in voice impressions is smaller than the variance in behavioral impressions. We found the effects of sentence meanings on voice impressions (Experiment 2) to be smaller than the effects of voice on behavioral impressions (Experiment 1). This might be due to the small variations in speakers. Future research should include more speakers to generate voice stimuli and examine the effects of sentence meanings on diverse impressions of voice. Third, in Experiment 2, participants were asked to ignore the emotional meanings of the sentences to exclude over-attention to sentence meaning. However, this instruction would be stronger than that of Experiment 1 and would weaken the meaning effects on voice impressions. Future research should confirm the true effect size of sentence meanings on voice impressions without this instruction. Finally, we only investigated impressions in a broad sense (from positive to negative) because the behavioral descriptions we used had been created previously, along with the dimensions [26], to examine first impression processing. However, social impressions could be differentiated based on several dimensions. Recent research suggests that across various domains, trustworthiness and dominance are the most prominent of perceived qualities [12, 44, 45]. Future research needs to verify which impressions are more affected by either prosody or sentences during the formation of impressions of spoken language.

5. CONCLUSION

The present study examined whether prosody modulated social impression formation by using DAVID, an advanced voice transformation system. The outcomes demonstrated that prosody modulated the social impressions—as described by spoken language—toward suppressing the original impressions. In addition, impressions of the speaker’s voice were also influenced by the affective properties of the sentence meaning, which was not directly related to the speaker. These findings imply that people integrate information about what a speaker says and how he/she says it during impression formation for both speakers and others in spoken communication.

ACKNOWLEDGEMENTS

This work was supported by Grant-in-Aid for Scientific Research (C) of Japan to KM (19K03386, 19KK0313), Scientific Research on Innovative Areas (17H06344), JST-Mirai program (20349063), and JST-CREST (16H17876) to KW.
REFERENCES

1. Olivola, C.Y., and Todorov, A.; Elected in 100 milliseconds: Appearance-based trait inferences and voting. Journal of Nonverbal Behavior, 34(2), pp.83-110, 2010.

2. Todorov, A., Pakrashi, M., and Oosterhof, N.N.; Evaluating faces on trustworthiness after minimal time exposure. Social Cognition, 27(6), pp.813-833, 2009.

3. Todorov, A., and Porter, J.M.; Misleading first impressions: Different for different facial images of the same person. Psychological Science, 25(7), pp.1404-1417, 2014.

4. Willis, J., and Todorov, A.; First impressions: Making up your mind after a 100-ms exposure to a face. Psychological Science, 17(7), pp.592-598, 2006.

5. Ames, D.L., Fiske, S.T., and Todorov, A.; Impression formations: A focus on others’ intents, The Oxford Handbook of Social Neuroscience, pp.419-433, 2011.

6. Dunbar, R.I.M., Marriott, A., and Duncan, N.D.C.; Human conversational behavior. Human Nature, 8(3), pp.231-246, 1997.

7. Gilovich, T.; Secondhand information and social judgment. Journal of Experimental Social Psychology, 23(1), pp.59-74, 1987.

8. Inman, M.L., McDonald, N., and Ruch, A.; Boasting and firsthand and secondhand impressions: A new explanation for the positive teller-listener extremity effect. Basic and Applied Social Psychology, 26, pp.59-75, 2004.

9. Collins, S.A.; Men’s voices and women’s choices. Animal Behaviour, 60(6), pp.773-780, 2000.

10. Collins, S.A., and Missing, C.; Vocal and visual attractiveness are related in women. Animal Behaviour, 65(5), pp.997-1004, 2003.

11. Vukovic, J., Jones, B.C., Feinberg, D.R., DeBruine, L.M., Smith, F.G., Welling, L.L.M., and Little, A.C.; Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women’s preferences for masculine pitch in men’s voices. British Journal of Psychology, 102(1), pp.37-48, 2011.

12. McAleer, P., Todorov, A., and Belin, P.; How do you say ‘Hello’? Personality impressions from brief novel voices. PloS one, 9(3), e90779, 2014.

13. Aucouturier, J.-J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., and Watanabe, K.; Covert digital manipulation of vocal emotion after speakers’ emotional states in a congruent direction. Proceedings of the National Academy of Sciences, 113(4), pp.948-953, 2016.

14. Ben-David, B.M., Multani, N., Shakuf, V., Rudzicz, F., and Van Lieshout, P.H.H.M.; Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. Journal of Speech, Language, and Hearing Research, 59(1), pp.72-89, 2016.

15. Jacob, H., Brück, C., Plewnia, C., Wildgruber, D.; Cerebral processing of prosodic emotional signals: Evaluation of a network model using rTMS. PloS one, 9, e105509, 2014.

16. Kotz, S.A., and Paulmann, S.; When emotional prosody and semantics dance cheek to cheek: ERP evidence. Brain Research, 1151, pp.107-118, 2007.

17. Paulmann, S., and Kotz, S.A.; An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. Brain and Language, 105(1), pp.59-69, 2008.

18. Ishii, K., Reyes, J.A., and Kitayama, S.; Spontaneous attention to word content versus emotional tone: Differences among three cultures. Psychological Science, 14(1) pp.39-46, 2003.

19. Kitayama, S., and Ishii, K.; Word and voice: Spontaneous attention to emotional utterances in two languages. Cognition & Emotion, 16(1), pp.29-59, 2002.

20. Nygaard, L.C., and Lunders, E.R.; Resolution of lexical ambiguity by emotional tone of voice. Memory & Cognition, 30(4), pp.583-593, 2002.

21. Nygaard, L.C., and Queen, J.S.; Communicating emotion: Linking affective prosody and word meaning. Journal of Experimental Psychology: Human Perception and Performance, 34(4), pp.1017-1030, 2008.

22. Pell, M.D., Jaywant, A., Monetta, L., and Kotz, S.A.; Emotional speech processing: Disentangling the effects of prosody and semantic cues. Cognition & Emotion, 25(5), pp.834-853, 2011.

23. Schwartz, R., and Pell, M.D.; Emotional speech processing at the intersection of prosody and semantics. PloS one, 7(10), e47279, 2012.

24. Rachman, L., Liuni, M., Arias, P., Lind, A., Johansson, P., Hall, L., Richardson, D., Watanabe, K., Dubal, S., and Aucouturier, J.-J.; DAVID: An open-source platform for real-time transformation of infra-segmental emotional cues in running speech. Behavior Research Methods, 50(1), pp.323-343, 2018.

25. Bestelmeyer, P.E.G., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., and Belin, P.; Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. Cerebral Cortex, 22(6), pp.1263-1270, 2012.
26. Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., and Phelps, E. A.; A neural mechanism of first impressions. Nature Neuroscience, 12(4), pp.508-514, 2009.

27. Asch, S. E.; Forming impressions of personality. The Journal of Abnormal and Social Psychology, 41(3), pp.258-290, 1946.

28. Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., and Lawrence, M.; Dispositional attribution: Multiple inferences about motive-related traits. Journal of Personality and Social Psychology, 86(4), pp.530-540, 2004.

29. Anderson, N. H.; Averaging versus adding as a stimulus-combination rule in impression formation. Journal of Experimental Psychology, 70(4), pp.394-400, 1965.

30. Roese, N. J., and Morris, M. W.; Impression valence constrains social explanations: The case of discounting versus conjunction effects. Journal of Personality and Social Psychology, 77(3), pp.437-448, 1999.

31. Skowronski, J. J., and Carlston, D. E.; Negativity and extremity biases in impression formation: A review of explanations. Psychological Bulletin, 105(1), pp.131-142, 1989.

32. Hamilton, D. L., and Sherman, S. J.; Perceiving Persons and Groups. Psychological Review, 103(2), pp.336-355, 1996.

33. Brainard, D. H.; The psychophysics toolbox. Spatial Vision, 10(4), pp.433-436, 1997.

34. Pelli, D. G.; The Video Toolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), pp.437-442, 1997.

35. Calbi, M., Heimann, K., Barratt, D., Siri, F., Umiltà, M.A., and Gallese, V.; How context influences our perception of emotional faces: A behavioral study on the Kuleshov effect. Frontiers in Psychology, 8, 1684, 2017.

36. Bachorowski, J.-A.; Vocal expression and perception of emotion. Current Directions in Psychological Science, 8(2), pp.53-57, 1999.

37. Van der Zanden, T., Mos, M. B. J., Schouten, A. P., and Krahmer, E. J.; What people look at in multimodal online dating profiles: How Pictorial and textual cues affect impression formation. Communication Research, 2021, doi:10.1177/0093650221995316

38. Mehrabian, A., and Wiener, M.; Decoding of inconsistent communications. Journal of Personality and Social Psychology, 6(1), pp.109-114, 1967.

39. Carroll, J. M., and Russell, J. A.; Do facial expressions signal specific emotions? Judging emotion from the face in context. Journal of Personality and Social Psychology, 70(2), pp.205-218, 1996.

40. Aviezer, H., Trope, Y., and Todorov, A.; Holistic person processing: Faces and bodies tell the whole story. Journal of Personality and Social Psychology, 103(1), pp.20-37, 2012.

41. Righart, R., and de Gelder, B.; Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. Social Cognitive and Affective Neuroscience, 3(3), pp.270-278, 2008.

42. de Gelder, B.; Towards the neurobiology of emotional body language. Nature Reviews Neuroscience, 7(3), pp.242-249, 2006.

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

44. Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burtd, D. M., and Younga, A. W.; Social inferences from faces: Ambient images generate a three-dimensional model. Cognition, 127(1), pp.105-118, 2013.

45. Oosterhof, N. N., and Todorov, A.; The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), pp.11087-11092, 2008.

Miho Kitamura (Member)
Miho Kitamura is researcher at Waseda Institute for Advanced Study (Waseda University) and IdeaLab Inc., Japan. Her interests are social communication and emotion.

Katsumi Watanabe (Non-member)
Katsumi Watanabe is Professor at Waseda University (Department of Intermedia Art and Science), Japan and at University of New South Wales (Arts, Design and Architecture), Australia. His interests are human perception, cognition, and the brain.
### SUPPLEMENTARY INFORMATION

A total of 54 sentences described specific behaviors. Each sentence belongs to one of the emotional categories (positive, neutral, or negative).

| Category | Behavior Sentences (In Japanese) | Behavior Sentences (In English) |
|----------|----------------------------------|----------------------------------|
| Neutral  | 畑を耕して野菜を収穫した。 | He cultivated his own vegetables. |
| Neutral  | ベースボールドームに入場した。 | He lost 9 kilograms on the diet. |
| Neutral  | 新聞を読んでクエスチョン紙を見て10分で終わらせた。 | He finished the New York Times crossword puzzle in only 10 minutes. |
| Neutral  | もっと私が勝つと思ってなかった人が私の祝福を受け入れた。 | He accepted the congratulations from those who said he would not win the competition. |
| Neutral  | 親友が音楽家と音楽家であることに気づき。 | He noted that his best friends were scholars and musicians. |
| Neutral  | 達成と連携を取り合うことを約束した。 | He promised to keep in touch with his friends. |
| Neutral  | 新人を自分の家に迎えた。 | He invited the newcomers to his house. |
| Neutral  | 餐を近くからお食い Voltageを錯した。 | He danced around the bathroom as he brushed his teeth. |
| Neutral  | パークでゴルフを始めた。 | He played the devil advocate, just for fun. |
| Neutral  | とこの農場で数日過ごした。 | He spent several days on his cousin’s farm. |
| Neutral  | 新しいティールを日食した。 | He received the local woods instead of going to the mall with his friends. |
| Neutral  | 地元の高校で数学と英語を教えていた。 | He tutored a local high school student in both math and English literature. |
| Neutral  | みんなと手をつなぎ調子はとうとう悪化した。 | He shook everyone's hands and asked them how they were. |
| Neutral  | 旅行の写真を100枚見せた。 | He showed 100 slides of his trip. |
| Neutral  | 達成が気になって彼の歴史の授業のノートを彼女に貸した。 | He loaned her his notes from the history lectures she missed. |
| Neutral  | 試合がしんなりと科学的電気運搬を達成しに話し合った。 | He told a friend that he thought mindreading might be scientifically supported someday. |
| Neutral  | パーティーで他人的会話を立ち聞きした。 | He was eavesdropping on different conversations at the party. |
| Neutral  | 恥ずかしがって会話を聞こうとした。 | He tried to hear the conversation at the next table. |
| Neutral  | 学生たちに自分があまり役に立たないと言った。 | He told the student that he just wasn’t smart enough. |
| Neutral  | 達成とその恋人の会話を眺めました。 | He eavesdropped during his roommate's conversation with his girlfriend. |
| Neutral  | 買いに通っているに聞く耳を傾けた。 | He listened to his neighbor's conversation through the thin door. |
| Neutral  | 日本文学の学年レポートを友人から貸した。 | He borrowed a term paper for his Japanese literature course. |
| Neutral  | 物の人は書店を覗き込むと言ってた。 | He said that everyone else had missed the main point. |
| Neutral  | 声に泣きながら彼女を怒らしめた。 | He played the devil's advocate, just for fun. |
| Negative | 私もまた2年半のスクリプトの1番を着た。 | He put on one of his suits, both of which were tan. |
| Negative | 同じものを手でかすめてみた。 | He lied to get an extension on a deadline. |
| Negative | 朝日を見つめるために夜を徹底的に守った。 | He forgot a medical excuse to get out of the final exam. |
| Negative | 音で大きな声で耳打ちした。 | He whispered loudly to his friends at the funeral. |
| Negative | 新しい従業員に気が入らないのではいないか心配した。 | He worried that the new employees wouldn’t like him. |
| Negative | 畑の家に誰が来て誰が去ったかをいちいち見ていた。 | He watched his neighbor's house to see who came and went. |
| Negative | 雑誌を購入するために会社に去り続けると電話をした。 | He called in sick in order to read magazines. |
| Negative | 彼は人から汚物を残して外した。 | He left his girlfriend with 20 pounds of laundry. |
| Negative | 彼の使命を記入すると手の中も勇気をかえた。 | He made half a dozen errors in filling out his document. |
| Negative | ガードレットに引かれて足許をくすぐった。 | He tripped on the rug and twisted his ankle. |
| Negative | インターフェースの最前線に隠しを忘れた。 | He forgot his clothes in the laundromat dryer. |
| Negative | 本から一部をコピーして自分で書いたのだと主張した。 | He copied a passage from a book and claimed that it was his own writing. |
| Negative | データをもう1度クライアントにひっ掻いた。 | He knocked the wine glass over while reaching for the cheese. |
| Negative | 電化を用いる効果を想定してしまったことに気がついた。 | He realized his wallet was in his other pants when he went to pay for his lunch. |
| Negative | カフェで飲んだコーヒを飲む前に千円札を取り出した。 | He took two thousand yen from the wallet he found in the cafeteria before returning it to the owner. |
| Negative | 行列に並んでいる老人の前に割り込んだ。 | He stepped in front of the old man in line. |
| Negative | 自転車に乗っていると旗を振りながら走った。 | While riding his bicycle, he hit a parked car. |
| Negative | エンターテイメントの当日チケットを家から引き出し忘れてしまった。 | He left the concert tickets on the drawer in his room. |