VOICE-BASED PERSON PERCEPTION: TWO DIMENSIONS AND THEIR PHONETIC PROPERTIES

Mihoko Teshigawara

Komazawa University, Japan
mteshi@komazawa-u.ac.jp

ABSTRACT

This paper proposes a two-dimensional model of voice-based person perception, building on previous findings from social cognition. These two dimensions (warmth or valence and competence or dominance) are compared with two dimensions of animal vocalizations (affective state and size). This two-dimensional model appears compatible with findings from previous studies on vocal stereotypes and visual face trait evaluation.

Keywords: voice-based person perception, social cognition, voice quality, frequency code, vocal stereotypes

1. INTRODUCTION

Previous studies on vocal stereotypes [4, 18], in which experiment participants listened to voices and rated personality and vocal characteristics, reveal that people infer similar personality traits from voices. However, many previous studies considered only a few primitive phonetic correlates or none at all, relying instead on the correlation between judges' ratings of personality traits and auditory-perceptual assessments of vocal characteristics (see Section 1.2.2 of [14] for a review of such studies). Exceptions include a series of studies conducted by Uchida [17], which systematically examines the relationship between a set of personality traits from the NEO Personality Inventory [7] and phonetic properties, such as the experimental manipulation of pitch contours and speech rate. However, it may be problematic how well such adjectival phrases as *heavy-laden*, *foresighted* and *deliberate*, trait items in Uchida's studies, can elicit listeners' impressions of the target voices and whether such traits are fundamental to the first impressions formed when encountering strangers in real life. Instead, more innate and fundamental dimensions should mediate voice-based person perception.

I propose a two-dimensional model of voice-based person perception, drawing on previous findings from social cognition [2] including facial perception [11]. Inspired by Ohala's research [10], these two dimensions are compared with two dimensions that are composites of the Motivation-Structural rules of animal vocalizations [8, 9], from which I formulate hypotheses about their phonetic properties and demonstrate how these two dimensions are comparable to previous findings on vocal stereotypes [15, 16].

2. TWO DIMENSIONS IN SOCIAL COGNITION

2.1. Universal dimensions of social cognition

According to Fiske et al. [2], recent research has established that perceived warmth (or trustworthiness, i.e., perceived intent) and competence (i.e., perceived ability) are two universal dimensions of human social cognition. That is, when encountering strangers, people immediately judge the intent of the other and whether the other person can act on their intention. In addition, Fiske et al. mention that warmth judgments are primary, that warmth judgment occurs before that of competence and is more important than the latter. They also note "the warmth dimension predicts the valence of the interpersonal judgment (i.e., whether the impression is positive or negative)" ([2], p. 78).

2.2. Face evaluation

Oosterhof and Todorov [11] use a data-driven approach to identify fundamental dimensions of trait inference from faces. In their study, two dimensions were first constructed from unconstrained descriptions of the target faces by performing a principal components analysis, rather than by eliciting judgments on a set of *a priori* assumed traits (e.g., trustworthiness and aggressiveness), as is often done in this type of research. The two components were interpreted as valence (or trustworthiness) and dominance: all positive judgments (e.g., *trustworthy*, *responsible*) had positive and all negative judgments (e.g., *weird*, *unhappy*) had negative loadings on the first

principal component (valence/trustworthiness), whereas judgments of dominance, aggressiveness, and confidence had the highest loading on the second principal component. As Oosterhof and Todorov themselves note, despite the difference in the approaches, the two dimensions that emerged in their research, that is, valence and dominance, are comparable to the two universal dimensions of social cognition, warmth and competence.

In subsequent experiments, Oosterhof and Todorov used 300 computer-generated emotionally neutral faces that varied in standard deviations along continua of the two dimensions. In one experiment, after finding that *happy* and *angry* received statistically significant responses among the six basic emotions, they asked participants to judge the faces on a 9-point scale ranging from 1 (angry) to 9 (happy). They also obtained 9-point scale ratings for facial masculinity-femininity and maturity-immaturity in subsequent experiments. Their findings include the following: (i) the evaluation of a valence/trustworthiness dimension is strongly related to happy-angry judgments (i.e., the more trustworthy a face was judged to be, the happier participants judged it) and (ii) judgments of facial masculinity and maturity were stronger for the dominance dimension than for the valence dimension, suggesting that the dominance dimension is sensitive to features signaling physical strength.

Considering that voice-based person perception is also a subcategory of social cognition, along with person perception based on faces, it would be reasonable to hypothesize that there are also two fundamental dimensions in voice-based person perception, comparable to those of Oosterhof and Todorov's, that is, warmth or valence and competence or dominance. In addition, the phonetic properties of the warmth/valence dimension may resemble features of affective (happy-angry) vocalizations, whereas those of the competence/dominance dimension may resemble features signaling physical strength (i.e., masculinity and maturity).

3. TWO DIMENSIONS OF ANIMAL VOCALIZATIONS

Studies on animal vocalizations may seem somewhat irrelevant to the current discussion. However, Ohala's now-established frequency code theory [10] integrates part of Morton's Motivation-Structural rules (M-S rules) for animal vocalizations [8, 9] and merits review. Morton widely reviewed prior research on vocal communication signals of birds and mammals and proposed the following Motivation-Structural rules: "birds and mammals use harsh, relatively low-frequency sounds when hostile and higher-frequency, more pure tone-like sounds when frightened, appeasing, or approaching in a friendly manner" ([8], p. 855). Drawing upon this version of the M-S rules, Ohala proposed the frequency code to explain the cross-species F_0-function correlation. Simply put, a high or rising F_0 is associated with smallness, non-threatening attitude, desire for goodwill from the receiver, etc., whereas a low or falling F_0 is associated with largeness, threat, self-confidence, etc. Referring to Morton's M-S rules, Ohala also makes a connection between facial expressions (e.g., a smile and its opposite, the "o-face") and the frequency code, suggesting that resonances (i.e., vowel formants) convey an impression of the size of the signaler. (But some studies have found formant dispersion to be a better predictor of the body size of the signaler than F_0, e.g., [3].) However, Ohala did not address Morton's findings on sound quality, limiting his discussion to F_0 and resonances. My proposal is compatible with both aspects of the M-S rules.

In his 1994 contribution to *Sound Symbolism*, [9], Morton modified the M-S rules slightly and separated them into two dimensions: the first dimension pertains to sound quality, ranging from broadband (harsh) to narrowband (tonal); the second, to fundamental frequency or spectral distribution of sound energy ([9], p. 356). Morton does not explicitly associate the first dimension with affective states, possibly because it is difficult to know what emotions an animal feels (cf. [9], p. 357). However, in comparing acoustic characteristics of call types of the squirrel monkey differing in aversion, by stimulating implanted intra-cerebral electrodes, Fichtel et al. [1] found an increase in the ratio of non-harmonic to harmonic energy for the more aversive calls, suggestive of the calls' harsher quality. Therefore, it might be said that the first dimension of the M-S rules (quality) could correspond to the affective state of the signaler. The second dimension is compatible with Ohala's frequency code – the correspondence between the size of the signaler and F_0 (and vowel formants).

If we compare the two dimensions of animal vocalizations with the two fundamental dimensions of voice-based person perception, the first

dimension of the M-S rules (i.e., quality) may be related to the animal's affective state [1]. In voice-based person perception, I note these similarities: the phonetic properties of the first dimension in voice-based person perception, i.e., the warmth/valence dimension, resemble features of affective (happy-angry) vocalizations. Therefore, it could be hypothesized that the phonetic correlates of the first dimension of voice-based person perception may be voice qualities varying with the speaker's affective state.

The second dimension of the M-S rules pertains to F_0 and spectral distribution, which is virtually Ohala's frequency code, a correspondence between the size of the signaler and F_0 (and vowel formants). The second dimension of voice-based person perception (i.e., competence/dominance) has been hypothesized to resemble features signaling physical strength (i.e., masculinity and maturity). As females and children are usually smaller than adult males, one could hypothesize that phonetic correlates of the second dimension of voice-based person perception may be F_0 and vowel formants.

I propose two fundamental dimensions in voice-based person perception:

1. A warmth/valence dimension, whose phonetic properties resemble features of affective (happy-angry) vocalizations, that is, voice quality.
2. A competence/dominance dimension, whose phonetic properties resemble features signaling physical strength (i.e., masculinity and maturity), that is, F_0 and vowel formants.

Section 4 compares these dimensions with studies by Teshigawara and colleagues [15, 16].

4. STUDIES OF *ANIME* VOCAL STEREOTYPES

Teshigawara [15] investigated phonetic properties of vocal stereotypes of good and bad characters in Japanese culture using voices of heroes and villains in Japanese *anime* (animated cartoons). Following an auditory analysis of heroes' and villains' voices, a subset of characters varying in degree of laryngeal constriction and/or larynx lowering was chosen as stimuli for a perceptual experiment and masked using the random-splicing technique [12], yielding 5-second speech excerpts of 27 target speakers. Participants recorded their impressions for each speaker using 19 descriptors (physical and personality traits, emotional states, and vocal characteristics) on a 7-point scale.

Pearson's correlations, performed on excerpts for each gender separately, revealed that most positive trait items were correlated with one another in the ratings. Also Pearson's correlations between trait ratings and select phonetic measures found that the degree of laryngeal constriction was negatively correlated with the largest number of favorable trait items (e.g., *good-looking, loyal, intelligent*). The more constricted the larynx was perceived to be, the less good-looking, loyal, intelligent, etc., the speaker was perceived to be. Laryngeal constriction is involved in the production of such voice qualities as harsh and creaky voice and is among the characteristics predicted to be present in production of voices with negative emotion [13]. Therefore, the negative correlation between the positive traits and degree of laryngeal constriction may correspond to the first dimension of warmth/valence hypothesized in the previous section.

In contrast, for female voice actors, the degree of breathiness was positively correlated with the same positive traits that had been negatively correlated with the degree of laryngeal constriction. The breathier the female voices were, the higher the ratings they received for favorable traits such as *selfless*. According to Scherer, breathiness is predicted for such positive emotional states as happiness [13], again supporting the hypothesis regarding the first dimension of voice-based person perception. However, it should also be noted that mean F_0, which should be a phonetic correlate of the second dimension, was also somewhat negatively correlated with fewer positive traits in Teshigawara's data.

This study also found a negative correlation pattern between mean F2 for /a/ and the physical characteristic *big* and the personality trait *strong*. However, mean F_0 did not significantly correlate with them (cf. [3]). These traits are comparable to the second dimension hypothesized for voice-based person perception, and mean F2 is also predicted as the phonetic correlate of this dimension. Therefore, it may be said that the correlations that emerge from this data support the second dimension I am proposing.

In addition, Teshigawara et al. [16] conducted another experiment with Hebrew listeners using the same stimuli, and the correlations between the trait items and phonetic measures were roughly the same as the Japanese listeners [15], suggesting some cross-cultural validity for these two dimensions in voice-based person perception.

5. FUTURE DIRECTIONS

In order to validate this proposed model of auditory aspects of social cognition with new data, I propose the following series of experiments, drawing on the paradigm of Oosterhof and Todorov's data-driven approach to face evaluation [11]. First, instead of using *a priori* assumed traits, unrestricted descriptions of 200 new target voices will be collected from listeners and subjected to a principal components analysis. New traits obtained from components with high loadings on them will then be used in subsequent experiments using a large number of synthesized voices in graded variation by means of Tandem-STRAIGHT [5]. Synthesized voices will vary in two dimensions: voice quality, including breathiness, harshness, and creakiness (Dimension 1); and F_0 and vowel formant frequencies (Dimension 2). These voices will be presented to listeners in order to evaluate vocal emotion and maturity (masculinity) as well as the newly obtained traits. Correlations between acoustic and auditory measures for the two dimensions and the trait ratings will be calculated in order to validate the proposed two dimensions.

This current paper can be compared to a face evaluation study that uses still photographs (e.g., [11]), rather than videography, in that it deals only with quasi-static aspects of voice (i.e., mean F_0, mean F2, voice quality as an overall tendency). Given that speech is inherently dynamic, temporal aspects, such as pitch contours (cf. [17]), should also be incorporated into phonetic measures. Despite these shortcomings, the current hypotheses are well worth validating with new data.

6. ACKNOWLEDGMENTS

I gratefully acknowledge support by KAKENHI Nos. 19320060, 19700173, & 23320087, MEXT, Japan.

7. REFERENCES

[1] Fichtel, C., Hammerschmidt, K., Jürgens, U. 2001. On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behaviour* 138, 97-116.

[2] Fiske, S.T., Cuddy, A.J.C., Glick, P. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends Cogn. Sci.* 11, 77-83.

[3] Fitch, W.T. 1997. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* 102, 1213-1222.

[4] Hecht, M.A., LaFrance, M. 1995. How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *J. Appl. Soc. Psychol.* 25, 2086-2098.

[5] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., Banno, H. 2008. Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *Proc. ICASSP*, 3933-3936.

[6] Laver, J. 1994. *Principles of Phonetics.* Cambridge: Cambridge University Press.

[7] McCrae, R.R., Costa, P.T. Jr. 1987. Validation of the five-factor model of personality across instruments and observers. *J. Pers. Soc. Psychol.* 52, 81-90.

[8] Morton, E.S. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111, 855-869.

[9] Morton, E.S. 1994. Sound symbolism and its role in non-human vertebrate communication. In Hinton, L., Nichols, J., Ohala, J.J. (eds.), *Sound Symbolism.* Cambridge: Cambridge University Press, 348-365.

[10] Ohala, J.J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In Hinton, L., Nichols, J., Ohala, J.J. (eds.), *Sound Symbolism*. Cambridge: Cambridge University Press, 325-347.

[11] Oosterhof, N.N., Todorov, A. 2008. The functional basis of face evaluation. *PNAS* 105, 32, 11087-11092.

[12] Scherer, K.R. 1971. Randomized splicing: A note on a simple technique for masking speech content. *J. Exp. Res. Pers.* 5, 155-159.

[13] Scherer, K.R. 1986. Vocal affect expression: A review and a model for future research. *Psychol. Bull.* 99, 143-165.

[14] Teshigawara, M. 2003. Voices in Japanese Animation: A Phonetic Study of Vocal Stereotypes of Heroes and Villains in Japanese Culture. Ph.D. dissertation, University of Victoria, Canada. *http://web.uvic.ca/ling/students/graduate/Dissertation_Teshigawara.pdf*

[15] Teshigawara, M. 2009. Vocal expressions of emotions and personalities in Japanese anime. In Izdebski, K. (ed.), *Emotions of the Human Voice, Vol. III Culture and Perception*. San Diego: Plural Publishing, 275-287.

[16] Teshigawara, M., Amir, N., Amir, O., Milano Wlosko, E., Avivi M. 2009. Perceptions of Japanese anime voices by Hebrew speakers. In Izdebski, K. (ed.), *Emotions of the Human Voice, Vol. III Culture and Perception*. San Diego: Plural Publishing, 189-198.

[17] Uchida, T. 2007. Effects of F_0 range and contours in speech upon the image of speakers' personality. *Proc. 19th ICA* Madrid. *http://www.sea-acustica.es/WEB_ICA_07/fchrs/papers/cas-03-024.pdf*

[18] Yarmey, A.D. 1993. Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Appl. Cognitive Psych*, 7, 419-431.