

Facial Expression and Vocal Pitch Height: Evidence of an Intermodal Association

DAVID HURON^[1]

School of Music, Ohio State University

SOFIA DAHL

Department of Media Technology, Aalborg University Copenhagen

RANDOLPH JOHNSON

School of Music, Ohio State University

ABSTRACT: Forty-four participants were asked to sing moderate, high, and low pitches while their faces were photographed. In a two-alternative forced choice task, independent judges selected the high-pitch faces as more friendly than the low-pitch faces. When photographs were cropped to show only the eye region, judges still rated the high-pitch faces friendlier than the low-pitch faces. These results are consistent with prior research showing that vocal pitch height is used to signal aggression (low pitch) or appeasement (high pitch). An analysis of the facial features shows a strong correlation between eyebrow position and sung pitch—consistent with the role of eyebrows in signaling aggression and appeasement. Overall, the results are consistent with an inter-modal linkage between vocal and facial expressions.

Submitted 2009 May 7; accepted 2009 July 8.

KEYWORDS: *pitch height, facial expression, aggression, intermodal perception*

IN animal behavior, size is commonly used to signal threat or appeasement: threat displays typically involve efforts to appear bigger, whereas appeasement displays typically involve efforts to appear smaller. Such displays may include both visual and auditory components. Large acoustical vibrators tend to have greater mass and so typically produce a lower fundamental frequency than smaller vibrators. In Charles Darwin's classic work on emotion (1872), he informally noted the role of pitch in emotional expression: Cebus monkeys produce a deep grunt as part of a threat display. By contrast, "a person complaining of ill treatment" he wrote, "almost always speaks in a high-pitched voice". That is, expressions of vulnerability employ a contrasting pitch compared with expressions of threat. Darwin proposed that these pitch differences are consistent with the larger/smaller postures used respectively in threat and appeasement displays. Morton (1977) expanded on Darwin's intuition by describing a more extensive set of animal observations. Morton explicitly proposed that low pitch symbolizes aggression throughout the animal kingdom due to its association with large size.

In the case of human speech, Morton, Bolinger, Ohala and others have established that fundamental vocal frequency (F0) is perceived to be associated with social cues for dominance and submissiveness. In a wide sample of cultures, Bolinger (1978) observed that high or rising vocal pitch is associated with politeness, deference, submissiveness, and lack of confidence. Conversely, low or falling vocal pitch is associated with authority, threat, aggression, and confidence. Ohala (1984) and Morton (1994) provided additional support for this association in non-human animals. Sound-size symbolism remains an active area of research (Ohala, 1994; Diffloth *et al.*, 2006; Tsur, 2006).

Apart from speech-related research, the association between pitch height, submissiveness, and dominance has also been demonstrated for musical stimuli. Huron, Kinney and Precoda (2006) played melodic phrases to listeners who variously judged the aggressiveness, threat, politeness, or submissiveness of each phrase (see also Williams & Stevens, 1972; Scherer *et al.*, 1973; Morton, 2006). A significant main effect was demonstrated where judgments were correlated with the overall pitch transposition of a given

melodic phrase. In general, the same melody played at a lower pitch height sounds more aggressive, more threatening, less polite, and less submissive.

In parallel with the research related to auditory displays, extensive research has been carried out related to the visual signaling of aggression and appeasement. This research has focused almost exclusively on facial expressions. In threat and appeasement displays, the two principal facial features are found in the mouth and eyebrow regions (Ekman, 1980, 2003). When angry, the eyebrows are lowered resulting in a more pronounced brow ridge (Eibl-Eibesfeldt, 1979). In most of the cultures sampled by Keating *et al.* (1981), lowered eyebrows were interpreted as displays of greater dominance or aggression. This effect is especially pronounced in males. Children have comparatively flat foreheads, but at puberty the brow ridges grow noticeably. Tanner (1978) observed that these ridges grow more appreciably in males compared with females. Conversely, raised or arched eyebrows are indicative of appeasement or friendliness. Again, sex differences have been observed. When at rest, mature males exhibit less arched eyebrows than mature females (Eibl-Eibesfeldt, 1979). Berry and McArthur (1986) have suggested that the application of cosmetics by women typically causes the face to be perceived as warmer and more submissive.

Appeasement and aggression signaling also produces significant changes in the mouth region. Flexing the zygomatic muscles causes an enlargement of the cheeks and also raises the corners of the mouth in the familiar smile. While most regard the smile as exclusively a visual signal, smiling also causes distinctive changes to the voice by reducing the volume of the oral cavity and shortening the effective length of the vocal tract. As a result, vocal formants rise upward in frequency (Tartter, 1980). This is evident in the so-called “smiling voice” which is often apparent when conversing on the telephone.

Interestingly, Ohala (1980) proposed that the smile originated as a vocal (rather than exclusively visual) display. Paradoxically, smiling has a tendency to expose the teeth, and the barring of teeth is commonly associated with threat rather than appeasement. However, Ohala suggested that the flexing of the zygomatic muscles may have been intended to reduce the volume of the oral cavity and so shift the vocal resonances upward in accordance with an auditory signal of appeasement. According to this view, the smile originated as a vocal display and was later generalized to include a visual aspect so that flexing the zygomatic muscles alone was sufficient to indicate appeasement or friendliness, even if no vocalization was produced (see also Ohala, 1982, 1983).

Ohala's theory of the origin of the smile suggests that one should observe a strong inter-modal link between visual and auditory elements involved in any threat or appeasement display. Evidence in support of coordinated facial and vocal displays has been demonstrated for chimpanzees. Bauer (1987) found a correlation between fundamental frequency in spontaneous vocalizations and teeth- and lip-opening distances. These observations lend support to the idea that appeasing facial displays are coordinated with appeasing vocalizations. However, no direct evidence of this relationship has been observed in humans.

One approach to exploring this presumed relationship is to ask participants to produce vocalizations of high and low pitches and examine the spontaneous concurrent facial expressions.

HYPOTHESES

We predict that singing of higher pitches will be associated with facial expressions linked to appeasement. Conversely, we predict that singing of low pitch will be associated with facial expressions linked to aggression.

SUBJECTS

For this experiment, a convenience sample of 44 subjects was recruited. Participants were passers-by on the campus of the Ohio State University a few days prior to the beginning of the academic year. At this time of year, the population of pedestrians includes a mix of students, parents, siblings, and university employees. As an inducement to participate, potential subjects were offered a free ice cream bar.

In order to satisfy Human Subjects restrictions, only participants 18 years of age and older were included. Otherwise, no effort was made to record the ages of the participants. In general, roughly equivalent numbers of males and females were evident in the population of passers-by. In total, 19 female and 25 male subjects participated in an experiment lasting less than one minute. No effort was made to collect further demographic or personal information. In particular, participants were not questioned regarding their level of musical involvement or musical expertise.

PROCEDURE

The experimenters wore broad-rimmed hats that covered their eyebrows and foreheads. A digital camera was mounted on a tripod at eye level (roughly 160 cm above the ground). Volunteers were asked to remove hats and sunglasses. In one case, a volunteer was asked to pull his hair back (so as not to obscure his eyebrows).

A recruitment sign was placed prominently near a busy pathway. The sign read: “OSU Research: A FREE ICE-CREAM BAR—for a song (and photo). We take three photographs of you while you sing a note. Photos are used only for research purposes and will not be distributed. Sorry: Must be 18 years or older. Study has been approved by the OSU Institutional Review Board.”

Instructions

When approached by a volunteer, the experimenter reiterated the procedure as posted on the sign. Experimental volunteers were prompted according to variants of the following script:

“I want you to sing a comfortable pitch and sustain it while we take your picture. Sing whatever vowel you like. Now hold the note . . . [take picture]. Now I want you to sing a higher [lower] pitch—the highest [lowest] pitch you can. Good. Now hold it.” [take picture]

The first photograph was always taken with the participant singing a pitch of their choosing. *A priori*, the first photograph was deemed by the experimenters to be a practice trial that would not be used in the analysis phase. The second and third photos involved singing the high and low pitches: the order of high/low pitch instructions was randomized for each subject. Although different vowels influence facial expression, the choice of vowel was nevertheless left up to the participant.

In order to identify each pitch condition, participants were asked to hold coded cards in their right hand near their chest while being photographed. Three cards contained geometric shapes to represent the three pitch conditions: a circle for the practice trial, a square for the low pitch, and a triangle for the high pitch. Participants were handed the cards in such a way that they rarely, if ever, saw the coded symbol. In addition to holding the card, participants also held a small electronic tuner in their left hand. The original goal of the tuner was to help the experimenters determine the pitch of the sung tone. However, the tuner information was discovered to be unreliable and so was not used for analysis.

Typically, participants appeared in groups of between two and five individuals. The social atmosphere was commonly cheerful, and participants were prone to smile or laugh. While being photographed, participants sometimes broke into laughter—especially if the participant was drawn from a group of companions. In order to minimize such social interactions, the camera position was physically arranged to make eye contact with their companions difficult.

In some cases, more than one photograph was taken for a single condition. For example, an additional photo might be taken if a person's hair dropped down to obscure the eyebrows or forehead, or if the participant stopped singing at the moment when the photo was taken. Of the 44 participants, 13 required an additional photograph. In 3 cases, 3 photographs were taken for a single condition. For the purposes of the analysis, only the final photo for each condition was used.

It should be noted that the very act of taking a non-clandestine picture of a person may lead to unwanted demand characteristics that can confound the experiment. Moreover, such demand characteristics could be aggravated by asking a participant to remove her sunglasses or pull his hair away from his face.

Rating of Facial Expressions

A simple test of the hypothesis is to compare pairs of photographs (for the high and low pitch conditions) in terms of relative friendliness. In order to avoid possible experimenter bias, we recruited four coders who were unfamiliar with the purpose of the experiment—two males and two females. The coders were predominantly graduate students in the School of Music. Each coder was shown pairs of photographs of the same individual and instructed to choose which facial expression is “friendlier.” All photographs were cropped and edited so that the coded card and pitch tuner was not visible. Photographs were presented as left-right pairs, and each coder received randomized pair orderings so that high and low photographs

appeared with roughly equal frequency in the left and right horizontal positions. In addition, each coder judged the paired photographs in a different random order.

Figure 1 shows a mock (experimenter-posed) example of high and low pitch faces. The left photo illustrates the singing of a low pitch, and the right photo illustrates the singing of a high pitch.



Figure 1. Experimenter-posed facial expressions associated with singing low pitches (left photo) and high pitches (right photo). Viewers rated high-pitched expressions as friendlier than low-pitched expressions.

In order to assess inter-rater reliability, the paired-photograph judgments were compared between pairs of coders. Table 1 compares the proportion of agreement between all four judges. If two judges judged all 44 paired photographs the same, then the agreement score would be 100 percent. A score of 50 percent represents a chance level of agreement.

	Judge 1	Judge 2	Judge 3	Judge 4
Judge 1	100			
Judge 2	90.9	100		
Judge 3	88.6	88.6	100	
Judge 4	68.2	63.6	61.4	100

Table 1. Inter-rater agreement for full-face photographs (in percent).

In general, the average agreement is 76.9 percent. One of the coders (Judge 4) appears to differ from the other coders. If the responses of Judge 4 are excluded, the average agreement is 89.4 percent. It may be noteworthy that Judge 4 was male; compared with females, males are known to be poorer judges of emotional expressions (Ekman, 2003). While the judgments are clearly not unanimous, they were deemed sufficiently consistent to warrant further analysis.

Pooling together all of the judgments, Table 2 shows the association between judged friendliness and pitch height for the 44 paired photographs. As can be seen, there is a significant association between sung pitch height and judged relative friendliness ($\chi^2 = 42.3$; $df = 1$; $p < 0.001$). *Prima facie*, this result is consistent with the hypothesis.

	More friendly
High pitch	149
Low pitch	27

Table 2. Association between judged friendliness and pitch height for 44 paired photographs.

A possible objection to these results is that the physical act of singing involves the mouth and neck region. It may simply be the case that singing higher pitches is easier if the zygomatic muscles around the mouth are flexed—resulting in a smile-like appearance. Similarly, singing low pitches may be easier if the chin is tilted downward, or the mouth is rounded.

As a possible control for this interpretation, we carried out a second analysis that focused on the region above the nose. As noted earlier, lowered eyebrows often signal aggression. For each original photograph, a cropped version was prepared in which the photo contained only the region above the nose tip.

In order to avoid possible experimenter bias, four new coders were recruited who were unfamiliar with the purpose of the experiment. The coders were predominantly graduate students or employees in the School of Music—two male and two female. Once again, each coder saw pairs of cropped photographs, and was instructed to identify which of the two photographs is “friendlier.”

Figure 2 shows a mock (experimenter-posed) example of the cropped photos. The left photo shows the singing of a low pitch, whereas the right photo shows the singing of a high pitch.



Figure 2. Cropped versions of the experimenter-posed photographs shown in Figure 1. All photos were cropped to show the face region above the nose tip. The left photo shows the singing of a low pitch, whereas the right photo shows the singing of a high pitch. Viewers again rated high-pitched expressions as friendlier than low-pitched expressions.

Table 3 shows the inter-rater agreement. Given the reduced visual information, we would expect the overall level of agreement to be lower than for the full-face photos—and that is indeed that case. The average inter-judge agreement is 67.0 percent. Once again, one of the coders (Judge 8) appears to differ from the other coders. If the responses of Judge 8 are excluded, the average agreement is 75.8 percent. It may be noteworthy that Judge 8 was also male.

	Judge 5	Judge 6	Judge 7	Judge 8
Judge 5	100			
Judge 6	70.5	100		
Judge 7	79.5	77.3	100	
Judge 8	61.4	54.5	59.1	100

Table 3. Inter-rater agreement for cropped photographs (in percent).

In general, the judgments show sufficient agreement to proceed with analysis of the pooled judgment data.

Table 4 shows the association between judged friendliness and pitch height for the 44 paired cropped photographs. As can be seen, there remains a significant association between sung pitch height and

relative friendliness ($\chi^2 = 19.1$; $df = 1$; $p < 0.001$). This result suggests that the relationship between vocalized pitch-height and facial expression is evident even when the upper region of the face is isolated. Once again, this result is consistent with the experimental hypothesis.

	More friendly
High pitch	129
Low pitch	47

Table 4. Association between judged friendliness and pitch height for 44 paired, cropped photographs.

DISCUSSION

A number of researchers have examined postural, kinematic, and facial expressions of musicians engaged in performance (see, for example, Dahl & Friberg, 2004, 2007; Thompson & Russo, 2006, 2007). With respect to the hypotheses tested here, pertinent qualitative research has been carried out by Bonfiglioli, Caterina, Incasa and Baroni (2006). Bonfiglioli *et al.* examined the eyebrow movements of famous pianists performing on video recordings. They related the eyebrow movements to musical features, including dynamics, articulation, pitch register, and other aspects. Based on descriptive non-quantitative observations, they reported that the players tend to raise their eyebrows when playing in medium and high pitch registers especially when performing at a quiet dynamic level. In addition, they observed that pianists tend to frown when performing at medium and low pitch registers at louder dynamic level. These observations are consistent with the results of the experiment reported here.

Overall, the analyses presented above suggest that there is an intermodal link between vocal and facial displays. That is, the results are consistent with the idea that the use of high pitch to signal appeasement or friendliness may be neurologically linked to appeasing or friendly facial expressions, and as a possible corollary, that the use of low pitch to signal aggression may be neurologically linked to aggressive facial expression.

Although the cause of this relationship is not the subject of this study, one might propose a number of speculative hypotheses regarding its origin. It may be that the vocal/facial link is a learned association that is wholly cultural in origin. More plausibly, the link may be a culturally-mediated exaggeration of a weak (though ubiquitous) size-sound symbolism. The relationship may be an artifact of an as yet undocumented anatomical or physiological link between the muscles of the face and the muscles of the laryngeal cavity. Another possibility is that the frequent combinations of low pitch/high pitch with aggressive/friendly expressions ultimately cause some sort of learned cross-motor connection. That is, initially independent motor control mechanisms for vocal height and facial expression become fused through frequent association. Such speculations suggest possible avenues for future research.

NOTES

[1] Please address all correspondence to: David Huron, School of Music, 1866 College Rd., Columbus, OH, 43210, U.S.A.

REFERENCES

- Bauer, H.R. (1987). Frequency code: Orofacial correlates of fundamental frequency. *Phonetica*, 44, 173-191.
- Berry, D.S., & McArthur, L.Z. (1986). Perceiving character in faces: The impact of age-related cranio-facial changes on social perception. *Psychological Bulletin*, 100, 3-18.
- Bolinger, D.L. (1978). Intonation across languages. In: J.H. Greenberg, C.A. Ferguson & E.A. Moravcsik (Eds.), *Universals of Human Language*, Vol. 2: *Phonology*. Stanford, CA: Stanford University Press, pp. 471-524.

- Bonfiglioli, L., Caterina, R., Incasa, I., & Baroni, M. (2006). Facial expression and piano performance. In: M. Baroni, A.R. Addressi, R. Caterina, M. Costa, *Proceedings of the 9th International Conference on Music Perception and Cognition*, pp. 1355-1360.
- Dahl, S., & Friberg, A. (2004). Expressiveness of musician's body movements in performances on marimba. In: A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction, Lectures Notes in Artificial Intelligence*, 2915, 549-556.
- Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, 24, 433-454.
- Darwin, C. (1872/1965). *Expression of the Emotions in Man and Animals*. Chicago: Chicago University Press.
- Diffloth, G. (1994). i: big, a: small. In: L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound Symbolism*. Cambridge: Cambridge University Press, pp. 107-114.
- Eibl-Eibesfeldt, I. (1979). *The Biology of Peace and War: Men, Animals, and Aggression*. New York: Viking.
- Ekman, P. (1980). *The Face of Man: Expressions of Universal Emotions in a New Guinea Village*. New York: Garland Press.
- Ekman, P. (2003). *Emotions Revealed: Understanding Faces and Feelings*. New York: Times Books.
- Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review*, 1(3): 170-177.
- Keating, C., Mazur, A., Segall, M., Cysneiros, p., Divale, W., Kilbride, J., Komin, S., Leahy, P., Thurman, B., & Wirsing, R. (1981). Culture and the Perception of Social Dominance from Facial Expression. *Journal of Personality and Social Psychology*, 40, 615-626.
- Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111, 855-869.
- Morton, E.S. (1994). Sound symbolism and its role in non-human vertebrate communication. In: L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound Symbolism*. Cambridge: Cambridge University Press, pp. 348-365.
- Morton, E.S. (2006). Commentary on "The influence of pitch height on the perception of submissiveness and threat in musical passages" by David Huron, Daryl Kinney, and Kristin Precoda. *Empirical Musicology Review*, 1(3): 178-179.
- Ohala, J. (1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America*, 68, S33.
- Ohala, J. (1982). The voice of dominance. *Journal of the Acoustical Society of America*, 72, S66.
- Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40, 1-18.
- Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica*, 41, 1-16.
- Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In: L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound Symbolism*. Cambridge: Cambridge University Press, pp. 325-347.

Scherer, K., London, H. & Wolf, J.J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, 7, 31-44.

Tanner, J.M. (1978). *Fetus into Man: Physical Growth from Conception to Maturity*. Cambridge, MA: Harvard University Press.

Tartter, V.C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, 27(1): 24-27.

Thompson, W.F., & Russo, F.A. (2006). Facial expressions of pitch structure. In: M. Baroni, A.R. Addessi, R. Caterina, M. Costa, *Proceedings of the 9th International Conference on Music Perception and Cognition*, pp. 1141-1143.

Thompson, W.F., & Russo, F.A. (2007). Facing the music. *Psychological Science*, 18(9): 756-757.

Tsur, R. (2006). Size-sound symbolism revisited. *Journal of Pragmatics*, 38(6): 905-924.

Williams, C.E., & Stevens, K.N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, 52(4): 1238-1250.