

# Facing the Music

**William Forde Thompson**

Macquarie University

**Frank A. Russo**

Ryerson University

[digital.library.ryerson.ca/object/216](https://digital.library.ryerson.ca/object/216)

Please Cite:

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

[doi:10.1111/j.1467-9280.2007.01973.x](https://doi.org/10.1111/j.1467-9280.2007.01973.x)



## Short Report

## Facing the Music

William Forde Thompson<sup>1</sup> and Frank A. Russo<sup>2</sup><sup>1</sup>Macquarie University, Sydney, Australia, and <sup>2</sup>Ryerson University, Toronto, Ontario, Canada

From the phonograph of the 19th century to the iPod today, music technologies have typically isolated the auditory dimension of music, filtering out nonacoustic information and transmitting what most people assume is the essence of music. Yet many esteemed performers over the past century, such as Judy Garland and B.B. King, are renowned for their dramatic use of facial expressions (Thompson, Graham, & Russo, 2005). Are such expressions merely show business, or are they integral to experiencing music?

In the investigation reported here, we considered whether the facial expressions and head movements of singers communicate melodic information that can be “read” by viewers. Three trained vocalists were recorded singing ascending melodic intervals. Subjects saw the visual recordings (without sound) and rated the size of the intervals they imagined the performers were singing.

## METHOD

Seventeen subjects (8 females, 11 males) from the University of Toronto community provided judgments. Subjects had 0 to 8 years of music training ( $M = 2.29$ ,  $SE = 0.58$ ) and ranged in age from 18 to 25 years ( $M = 19.71$ ,  $SE = 0.54$ ).

Three trained female vocalists sang 13 ascending melodic intervals spanning 0 to 12 semitones. Before attempting each interval, the vocalists heard the target interval through headphones. They then attempted to match the pitch and timing of the interval, articulating the syllable “la” on each note. The mean pitch height of intervals was centered on the middle of each singer’s range, and tones were 1.5 s in duration. The singers were instructed to sing naturally but accurately, and were not informed of the purpose of the experiment. Performances were recorded using an AKG C480B microphone and a Canon ZR60 digital camcorder and edited using Final Cut software on a Macintosh G4 computer. All sung intervals were within 1 cent of the intended interval size.

On the basis of the visual information, subjects rated interval size on a scale from 1 (*small interval*) to 7 (*large interval*).

---

Address correspondence to Bill Thompson, Department of Psychology, Macquarie University, Sydney, Australia, 2109, e-mail: bill.thompson@mq.edu.au.

## RESULTS

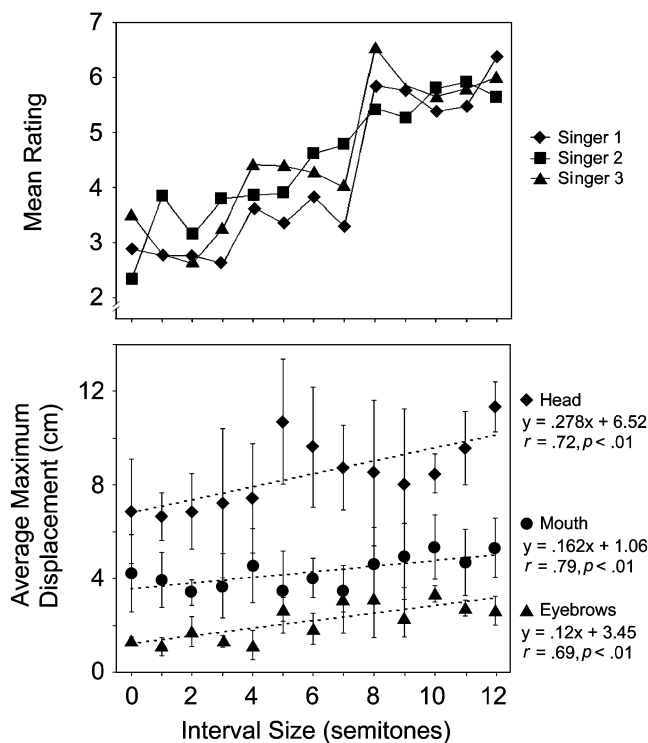
We conducted an analysis of covariance with musical training entered as a covariate. Ratings varied with veridical interval size,  $F(12, 180) = 15.8$ ,  $p < .0001$ , and musical training was not a significant covariate,  $F(1, 15) < 1$ . There was a significant interval-by-singer interaction,  $F(24, 360) = 2.28$ ,  $p < .001$ , but a high and significant correlation between interval size and mean ratings for all singers,  $r(11) = .90$ ,  $.95$ , and  $.88$  for Singers 1, 2, and 3, respectively ( $ps < .01$ ; see Fig. 1, top panel).

We used video-based motion tracking (Rockey, 2006) to calculate the maximum displacement of the head, eyebrows, and lips for each sung interval. We tracked the  $x$  and  $y$  screen coordinates (in pixels) of passive markers placed on the nose (for head movement), the top of each eyebrow, the upper edge of the upper lip, and the lower edge of the lower lips at a rate of 1 sample/frame (30 samples/s). For each frame, the Cartesian distance was computed for the head relative to its start position, the eyebrows relative to their start positions, and the lip opening. For each interval sung by each vocalist, the maxima of these distances were obtained and converted to real displacement values in centimeters. The bottom panel of Figure 1 shows the displacement values averaged across singers and confirms that the size of sung intervals was correlated with the degree of movement for all three features. The correlations suggest that the ratings of interval size were based on one or more of these cues.

We evaluated the hypothesis that visual information arising from our measures of movement mediated ratings of interval size. Interval size was significantly correlated with mean ratings across singers,  $r(11) = .94$ ,  $p < .001$  (Fig. 1, top panel), and with the movement measures,  $R = .94$ ,  $F(3, 9) = 24.22$ ,  $p < .001$  (Fig. 1, bottom panel). However, whereas interval size predicted 88% of the variance in mean ratings on its own, it accounted for less than 5% of the variance in mean ratings when the movement measures were statistically controlled,  $r_{sp}^2 = .049$ ,  $p > .05$  (Baron & Kenny, 1986).

## DISCUSSION

Subjects differentiated the size of sung intervals on the basis of facial expressions and head movements in the absence of an



**Fig. 1.** Mean ratings of interval size for the three singers (top panel) and maximum displacement of the head, mouth, and eyebrows as a function of interval size, averaged across the singers (bottom panel). The bottom panel also presents the regression line and correlation between displacement and interval size for each movement measure.

auditory signal. These results indicate that facial expressions carry information about pitch relations that can be read by viewers. Facial expressions and head movement may reflect interval size for several reasons. First, performers might directly communicate pitch relations through conscious or unconscious movements of their facial features and head. By mapping the extent of pitch change onto visually available movement information, performers can reinforce the size of an interval and facilitate melodic processing. Meaningful movements may also convey to listeners that pitch changes are intentional and interpretable. Second, performers may inadvertently move their facial features and head in response to an arousal state associated with pitch movement. Scherer (2003) observed that increased vocal pitch range is associated with heightened arousal. Thus, singing a large pitch interval may connote heightened

arousal that, in turn, is reflected in facial expressions and head movement. Third, adjustments in facial expression and head position may reflect facial and bodily movements introduced to optimize vocal production. For example, facial expressions used while singing a large ascending interval may reflect motor and attentional constraints associated with contraction of the cricothyroid muscle (Titze, 1994). Changing pitch while singing requires rapidly repositioning the vocal apparatus, and larger changes in pitch require greater degrees of repositioning.

A further possibility concerns the stapedius reflex, which controls the intensity of auditory input. The stapedius reflex is mediated by a neural network involving afferent input from the auditory nerve and efferent output to the facial nerve. In response to auditory feedback from self-vocalization, the stapedius muscle contracts, pulling the stapes of the middle ear away from the oval window. In addition, the tympani muscle contracts, pulling the malleus away from the eardrum. These contractions decrease transmission of energy at the cochlea. Although the degree of muscular activity engaged by the stapedius reflex is unlikely to induce significant facial and head movements, it may influence a performer's choice of expressive movements through a form of priming. That is, expressive singing may involve an elaboration of naturally occurring muscular activity.

## REFERENCES

- Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychology research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- Rockeby, D. (2006). softVNS 2.19g [Computer software]. Retrieved January 11, 2006, from <https://order.kagi.com/cgi-bin/store.cgi?storeID=K5D&&>
- Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227–256.
- Thompson, W.F., Graham, P., & Russo, F.A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 156*, 203–227.
- Titze, I.R. (1994). *Principles in voice production*. Englewood Cliffs, NJ: Prentice Hall.

(RECEIVED 8/7/06; REVISION ACCEPTED 12/12/06;  
FINAL MATERIALS RECEIVED 3/21/07)