Does Singing a Low-Pitch Tone Make You Look Angrier?

Peter Ahrendt, Christian C. Bach
Aalborg University
Master Program in Sound and Music Computing
{pahren16; cbach12}@student.aau.dk

Sofia Dahl
Aalborg University Copenhagen
Department of Architecture, Design
and Media Technology
sof@create.aau.dk

ABSTRACT

While many studies have shown that auditory and visual information influence each other, the link between some intermodal associations are less clear. We here replicate and extend an earlier experiment with ratings of pictures of people singing high and low-pitched tones. To this aim, we video recorded 19 participants singing high and low pitches and combined these into picture pairs. In a two-alternative forced choice test, two groups of six assessors were then asked to view the 19 picture pairs and select the "friendlier", and "angrier" expression respectively. The result is that assessors chose the high-pitch picture when they were asked to rate "friendlier" expression. Asking about "angrier" expression resulted in choosing the low-pitch picture. A non significant positive correlation between sung pitch ranges from every participant to the number of chosen high-pitch resp. low-pitch pictures was found.

1. INTRODUCTION

Music and speech research often departs from the auditory signal, but we also know that human communication is much more than what meets the ear. For music, evidence have accumulated that support the view of music as multimodal, where other modalities interact with and affect the communication. This has for example been shown in studies on musical communication of expression [1, 2], emotion [3], structure [4], perception of sung interval size [5,6] and note duration [7].

Apart from musical contexts, auditory information is also contributing to our percept of a person's mood and facial expression. When listening to a speaker, we can hear if the person is smiling [8]. Ohala proposed that the origin of the smile (which involves teeth showing and ought to be perceived as threat) evolved as an acoustic rather than a visual signal [9]. In brief, Ohala suggested that the high-pitch resonances produced by stretching the lips in a smile helps to make an acoustic display of appeasement by appearing smaller and non threatening (see [10] for an overview).

Huron and colleagues investigated the intermodal association between pitch and emotion in two studies [11, 12]. In a first study, 44 participants were asked to sing neutral, high and low notes in a vowel of their own choice while their photograph was taken. Independent assessors were then asked to look at high and low pitch picture pairs and select "the more friendlier looking". The high pitch pictures were chosen significantly more often, also when the pictures were cropped to only show the eye-region [11]. A subsequent study [12] found the reverse association in the measured vocal pitch height (f0) for speakers reading different vowels with eyebrows raised or lowered. The effect size was small, but significant.

While these two studies mentioned above show the intermodal associations from two perspectives, the first study [11] did not provide any data of the actual produced pitches. It would be reasonable to assume that the range of pitches produced by a singer would affect the changes in facial expression and thereby also an assessors selection. Moreover, ratings for the "opposite" (negative) expression Anger would bring further support to the intermodal connection vocal pitch and facial expression.

The objective of this study was twofold: 1) To replicate the experiment done in [11] but with additional assessment of "angrier" expressions. 2) To investigate whether the actual produced pitches influenced the choices of assessors. To this end we asked participants to produce three pitches while being recorded on video. Images and produced pitches were extracted from the videos and later two groups of assessors selected the "angrier" or "friendlier" looking face in a forced choice test.

We hypothesize that faces singing low-pitched tones will be selected as "angrier" looking to a higher degree than high-pitch singing faces, while faces singing high-pitched tones will be selected as "friendlier" more often than low-pitch faces. Furthermore we hypothesize that the relative distance between high and low notes will influence the expressions of singers, so that pitch-pair pictures for comparably large pitch ranges will be more often identified according to the first hypothesis than faces of singers producing smaller pitch ranges.

2. DATA COLLECTION

2.1 Participants

For the picture retrieval part, 19 test participants were recruited immediately after having completed an approxi-

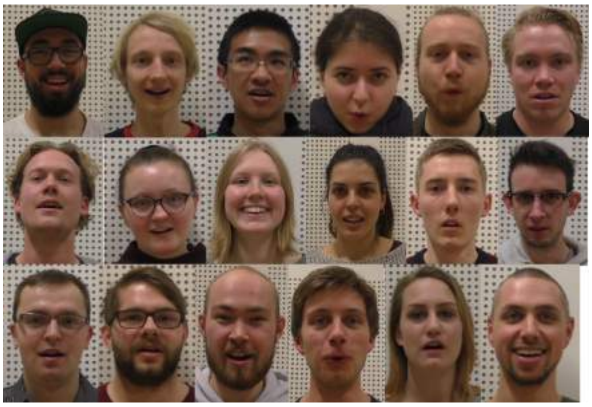

Figure 1. Some of the test persons participating in the test

mately 20 minutes long listening test. Participants were asked if they would like to be part of another experiment, involving them singing tones while recordings would be taken of them. While being a convenience sample, this way of recruiting also can be assumed to ensure that their emotional state would be relatively balanced as the test took place in a neutral testing environment. After being informed of the procedure, the volunteering participants gave their informed consent to being filmed. None of the participants received compensation for their participation.

2.2 Recording procedure

The environment was the Sound Lab of the Department of Architecture, Design and Media Technology, Aalborg University in Aalborg. The room has sound absorbing walls and no windows. The actual recordings took place in the main room which was divided by a wall with a window from an entrance room.

The participants were asked to stand in front of a regular camcorder placed on a tripod, and instructed to first produce a neutral effortless pitch, and then either a high followed by a low, or the reverse. High and low were in relation to their first pitch and the high-low order was determined using controlled randomization. The participants were instructed by variations of the following script presented by the experimenter:

We would like you to sing or produce a neutral tone which you are most comfortable with. After that tone please sing or produce a higher (lower) tone based on the neutral tone. After that the lower (higher) tone. You can sing whatever vowel you like. Please hold the tone a few seconds. Do you have any questions? [start recording] Now please sing the neutral tone. Good. Now the higher (lower) tone. Good. Now the lower (higher) tone. Good. [end recording]

To avoid that the situation would become stressful, which could have an influence on the facial expressions, we kept the task very simple. Thus, both vowel and pitches was left up to the participants (as done in [11]). Neither was any check for the actual vocal pitch range for each participant made.

While the participants were singing, the experimenter was standing behind the camera and focused on making the recording. That way participants would tend to look into the camera and not concentrate on the face of the experimenter. However, the participants were, not in all cases, asked to lock their eyes on the camera. Some of the participants wore glasses, which potentially could mask or interfere with the facial expression by covering, for example, the eyebrows. Well aware of the risk that less information would be conveyed, resulting in an increased difficulty for the assessors in the rating experiment, we did not ask the glasses to be removed.

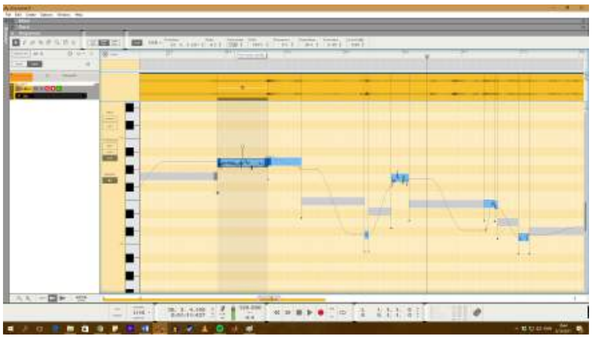

Figure 2. Tracking the pitch from the audio channel of the recorded video using the software Propellerhead Reason 9

The audio of the recorded video was exported into a DAW software called Reason 9 [13] to track and note the absolute pitches they produced in cents, MIDI notes and actual notes. In some cases the test participants were not able to sustain one clear note, so the average or the most prevalent pitch during each recording were chosen. The corresponding picture was then extracted from the video footage based on the time the pitch was produced, which would range from the first onset of the pitch produced (high or low) to the end. We did not settle for a fixed frame number across participants as they sometimes would blink, move around or in other ways reduce the clarity of the picture. Participants held, in most cases, tones three to four seconds long. In any case the picture chosen is one directly related to the pitch produced. The person who selected the pictures was not blind to the studies aim.

3. RATING EXPERIMENT

3.1 Participants

Two groups of six people from the local university campus volunteered as assessors to act in the rating experiment. The assessors were students who were currently in the building and approached by the experimenters. The rating of the pictures was made on a laptop on the site of the recruitment (usually open group working areas). None of the assessors had been involved in the first part of the test. The assessors received no compensation.

3.2 Stimuli and procedure

The produced images from the videos were assembled into pairs showing a high and low note per participant. A custom made testing interface made in the Matlab app designer [14] presented the picture pairs in a random order. For each picture pair, the program prompted the assessor to

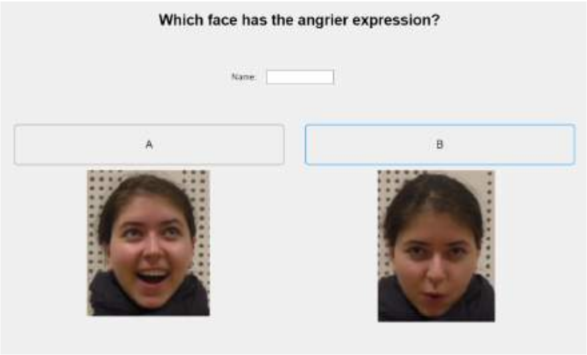
Which face has the angrier expression?

Figure 3. Test interface for choosing the face with the “angrier” expression

select either the more “angrier/friendlier” expression (one for each of the two groups of raters respectively) in a two-alternative forced choice design (see Figure 3). After a decision was made, the next picture pair of high/low-note singing faces appeared and a new choice was prompted. The assessors were not able to revise there choice. The experiment took about three minutes.

4. RESULTS

4.1 Sung pitches

Figure 4 shows a boxplot for the sung high-pitch and low-pitch tones (calculated in MIDI notes). The red lines inside the box represent the median values of the midi notes whereas the lower edges of the boxes depict the first quartile (25 % of the data). The upper edges represent the location of the third quartile (75 % of the data). The data enclosed in the boxes shows 50 % of the whole data sets. The whiskers of the boxplot illustrate in our case outliers which are below of the first quartile or above the third one. Additionally gray lines were plotted which indicate individual tone pairs of each participant. The lines start from the corresponding midi note of the high-pitch tone and end at the corresponding midi note of the low-pitch tone from the same participant. This shall give an overview of the pitch ranges produced during the experiment. The pitch ranges differ partially heavily among the participants. One participant for example sang a high pitch of approx. 3900 cents (108 in MIDI notes) (corresponding to C4) and a low pitch of approx. 900 cents (78 in MIDI notes) (corresponding to F#1), whereas another participant sang a high pitch note of approx. 1300 cents (82 in MIDI notes) (corresponding to A#1) and a low pitch of approx. 1200 cents (81 in MIDI notes) (corresponding to A1). A typical range was sung for example from approx. 2500 cents (94 in MIDI notes) (corresponding to A#1) to approx. 1000 cents (79 in MIDI notes) (corresponding to G1).

As can be seen, the ranges partially overlap, which is to be expected given the different vocal ranges of males and females. 14 of the participants were males, five females. We made no difference in analyzing data from male and female participants. The mean low pitch sung was at 1339 cents (82 in MIDI notes) (corresponding to A#1) and the mean high pitch at 2490 (corresponding to A#2) (94 in MIDI notes). The overall range of the recorded pitches spanned 3244 cents between the lowest (corresponding to D#1) and highest (C4) pitches recorded, with an average of 1151 cents, which is slightly below an octave.

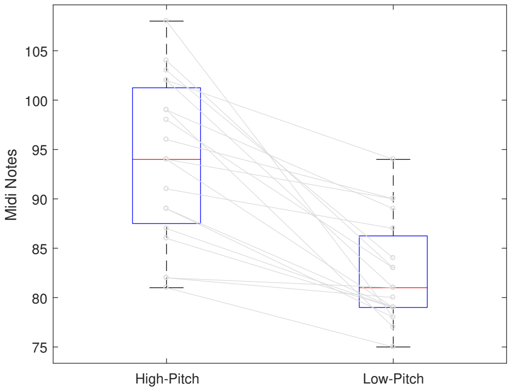

Figure 4. Boxplot for sung high- and low-pitch tones displayed in MIDI notes with individual tone pairs for each participant depicted with gray lines

	As.1	As. 2	As. 3	As. 4	As. 5	As. 6
As. 1	100					
As. 2	73.6	100				
As. 3	84.2	68.4	100			
As. 4	84.2	89.4	78.9	100		
As. 5	89.4	73.6	84.2	84.2	100	
As. 6	63.1	68.4	57.8	68.2	52.6	100

Table 1. “Angrier” inter-assessor agreement percentages

Please note that the depicted pitch ranges correspond to the pitch range the participants produced during the experiment. The participants were not tested in the maximum vocal pitch range they were able to produce.

4.2 Inter-Assessor Agreement

Table 1 and 2 show the inter-assessor agreement.

We calculated the inter-assessor agreement scores by summation of the agreeing choices between pairs of assessors, dividing it by the number of pictures. Hence, the scores do not indicate whether they chose the right picture or not, but whether two assessors selected the same picture. This was performed across all the pictures and assessor-pairs. If both assessors rated a given picture the same way, it counts towards the agreement aggregate.

Average assessor agreement was 74.6 % for selecting the face with an “angrier” expression. Assessor six seemed to differ from the others with an average of 62 %, at least 12 % below the rest. If assessor six is left out, the average agreement increases to 81 %. Table 2 shows the assessor agreements for selecting the “friendlier” face. For this group, there appears to be a higher agreement, with average scores between 73.64 % and 82.08 %, with a total average of 77.15 %.

	As.7	As. 8	As. 9	As. 10	As. 11	As. 12
As. 7	100					
As. 8	68.4	100				
As. 9	73.6	73.6	100			
As. 10	84.2	63.1	78.9	100		
As. 11	73.6	84.2	78.9	68.4	100	
As. 12	78.9	78.9	84.2	84.2	84.2	100

Table 2. “Friendlier” inter-assessor agreement percentages

Face	χ^2	df	p-value	High-Pitch	Low-Pitch
Angrier	21.93	1	$p < 0.01$	32	82
Friendlier	31.58	1	$p < 0.01$	87	27

Table 3. Associations between expression and number of high-pitch and low-pitch choices

Satisfied with the over all inter-assessor agreements for the two groups of assessors, we proceeded with the rest of the analysis and hypothesis testing.

4.3 Assessment of expression

The results of the choices of the assessors for the two experiments “angrier” and “friendlier” are summarized in table 3. As can be seen, the group assessing “angrier” predominantly selected the low-pitch faces (82 choices) compared to the high pitch faces (32 choices). Conversely, the other group of raters predominantly chose high-pitch faces as “friendlier” (87 choices) compared to low-pitch faces (27 choices).

Two Chi-square tests [15] confirmed that the sung pitches had a significant effect on the selected number of “angrier” and “friendlier” faces from the two groups of assessors. The χ^2 statistics from the test (calculated using the statistical software R) are also seen in Table 3.

The assessors’ picture selections support the hypothesis that facial expressions will give different emotional associations depending on the pitch sung. We now turn to the question whether a comparably large distance between high and low sung pitches would influence the facial expressions, thereby affecting the choises of assessors.

4.4 Correlation between Pitch range and Selections

To investigate whether there was a relationship between the difference between sung pitches and assessors’ choices, we calculated the produced *pitch range* in cents and compared this with the number of chosen high-pitch pictures for the “friendlier” facial expression and low-pitch pictures for “angrier” one respectively. The pitch range was calculated from the lowest to the highest pitch a participant sang. The relationship between the pitch range and the number of low-pitch faces chosen can be seen in Figure 5, while Figure 6 shows the relationship with number of high-pitch faces selected. As can be seen by the plotted trends there appear to be a weak positive relationship so that with an increase of pitch range the number of chosen pictures increases.

Face	r	t	df	p-value
Angrier	0.39	1.73	17	0.10
Friendlier	0.34	1.48	17	0.16

Table 4. Pearson’s product-moment correlation for number of chosen low-pitch (“angrier”) or high-pitch (“friendlier”) pictures and pitch range

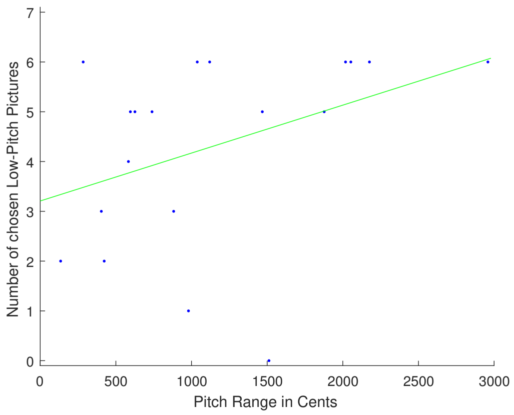

Figure 5. Correlation between the pitch range and number of low-pitch faces chosen. A linear regression curve is plotted in green.

A distribution investigation (Shapiro-Wilk normality test) resulted in the conclusion that the data for chosen number of pictures in each case was normally distributed while data for pitch range was not.

The statistics for two Pearson’s product-moment correlation [16] tests are presented in Table 4. The analysis showed no significant relationship between the range of pitches participants sang and the number of chosen high-pitch and low-pitch faces, respectively. Correlation coefficients were 0.39 and 0.34 for the selected low and high notes pictures respectively.

An outlier with a pitch range of 1512 cents can be seen in both Figures 5 and 6 where the number of chosen pictures for either high-pitch pictures or low-pitch pictures is zero. This means that two different groups of assessors rated the same pair of pictures of the same participant completely the opposite of what was predicted in our hypothesis. Notably the pitch range produced by this singer was near the average pitch range.

5. DISCUSSION

The results from our rating experiment showed that the high-pitched faces were selected as more friendlier than the low-pitch faces, replicating the results in [11]. In addition, our results also show the reverse relationship, with low-pitched faces selected as more angrier in comparison with the high-pitched faces. Our findings bring further support to intermodal associations between facial expression and vocal pitch height.

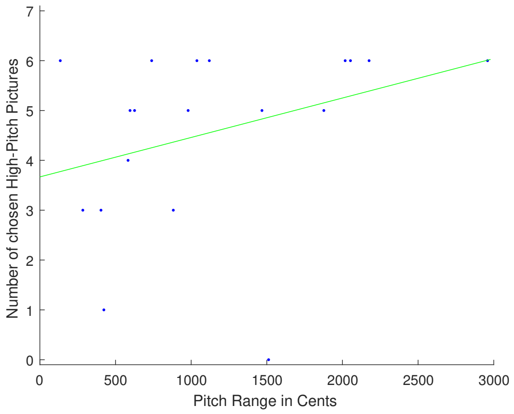

Figure 6. Correlation between the pitch range and number of high-pitch faces chosen. A linear regression curve is plotted in green.

A possible explanation for such intermodal associations could be the usefulness of conveying different perceived size (and hence threat) to other individuals (see e.g. [10, 12, 17]). In every day communication we typically shift between appeasing and threatening displays. For instance, infant-directed speech typically presents a non-threatening display with high pitch and raised eyebrows [18] while the infants older siebling might get told off by the same parent now frowning and using lower pitch in a more threatening display. Since parents come in many sizes with different vocal pitch ranges, it seems reasonable to assume that a combination of visual (raised or lower eyebrows) and vocal (high or low pitch) information would make the display less ambiguous to the receiver.

The reason for the eyebrow position in giving a friendly, non threatening, or an angry, more threatening display is subject to speculation. In a commentary to Huron et al. Ohala [17] suggested that raised eyebrows give the impression of proportionally larger eyes – something characteristic for newborns and infants. Since small children have relatively large eyes and are non threatening, giving the impression of larger eyes would also give the association of being smaller, less mature and non threatening. Conversely, the impression of smaller eyes would be associated with being more mature (and so possibly more of a threat) [17].

We hypothesized that the difference in pitch height, rather than the actual pitches per se, would influence the strength of the communication to the assessors. However the actual range in pitches produced by the singers in our study appeared to be of less importance, disproving our second hypothesis. The Pearson correlation between produced pitch range (in cents) and pitch-pictures selected by assessors, was 0.39 and 0.34 for the selected low and high notes pictures respectively. Although fairly small and non significant, the positive correlation is in line with the results of Thompson and Russo [5], who reported high correlation for observers rating sung interval size based on visual video information only. Thompson and Russo also measured the change in position of markers on the head, mouth, and eyebrows of the trained singers and found a positive correlation with sung interval size. Despite non-uniform data and using non-expert singers as participants in our study we find indication of similar relationship.

Some methodological choices in our study deserve further discussion. Firstly, the task of the singing participants was kept simple and both the vowel and pitches to sing was left to the individual (similar to the original study [11]). Since participants were not explicitly asked to maintain the first vowel they chose when singing the second note, three out of 19 participants shifted to a different vowel for the second note (see e.g. Figure 3). Different vowels affect the facial expressions and it is possible that this influenced the assessors choices. However, the same “free” instructions were used in the study by Huron et al. who found a significant effect also when only the eye-region was shown to assessors [11].

If the facial expressions associated with low and high pitches are part of a more general intermodal association, choosing different vowel could also be part of the signalling system. Ohala proposed that smiling helps to shape the acoustic signal to give a “friendlier”, more appeasing impression by shaping the filter to boost the higher frequencies [9]. As noted by Huron “the ‘opposite’ of the smile may not be the frown, but the pout” [10], something that has been observed as threat signal in many animal species. Pouting would indicate the production of rounded vowels that help to lengthen the vocal tract and produce lower resonances. In our experiment we counted nine participants who were pouting. Most of the pouting pictures were for the low pitch but in two cases participants also pouted with the high pitch, one of these participants produced relatively low pitches in general.

All singers started by producing a neutral, comfortable, tone, and in relation to this tone the higher and lower pitches were freely selected by the singers. As can be seen in Figure 4, the actual produced pitch ranges differed considerably. Some singers produced large differences while others chose a more modest range. If the facial expressions are related to the high and low pitches relative to each person’s individual vocal pitch span, an assessment of this one would have to be made. One way to address this in future studies could be to let a professional voice expert blindly assess how strained the voice appears for the produced notes.

A curious case is the specific set of picture pairs from one of the singers in our study that appeared as an outlier in Figures 5 and 6. Although the singer (Danish, male, had little facial hair, no glasses) produced notes with a reasonable interval range, the assessors consistently chose the low-note picture for the “friendlier” selection and the high-note picture for the “angrier”. Both pictures show the participant smiling with his head slightly turned, but the smile in the high-pitch picture may be described as more of a smirk while the singer is looking in a different direction. The low-pitch picture, on the other hand shows the participant looking straight into the camera. This could be a possible explanation that the low-pitch face appeared to present someone more approachable and friendly.

Apart from eyebrow height, there are other facial cues that observers use to judge what type of behaviour or personality an individual is likely to have. For instance, different facial morphology in both humans and chimpanzees reveal personality traits to observers [19] and an association between the width-to-height ratio and aggression ratings of static images showing neutral faces has been reported [20]. In our study we showed within-participant pairs of pictures to be compared by the assessors, but still some of these faces could have provided strong static cues not associated with the task of singing high and low notes. By rating the aggressiveness or friendlieness of individual pictures, possibly combined with eye tracking, it would be possible to better identify the specific role of the eyebrows in the assessment of the faces.

6. CONCLUSION

When singing notes of different pitch height, untrained singers' facial expressions are altered in such a way that when assessors are asked to select the "friendlier" face, they would choose the high-pitch picture to a higher degree over the low-pitch picture [11]. We could replicate this result from our experiment. Furthermore, we could also show the opposite relation where assessors instructed to select the "angrier" face, to a higher degree choose the low-tone picture over the high-pitch one. Hence, a task normally associated with music appear to affect dynamic facial expressions in a way that observers detect as emotional expressions influencing their choices of "angrier" and "friendlier" above level of chance.

Acknowledgments

The authors would like to thank all singing participants and assessors.
Authors Ahrendt and Bach were responsible for the main part of the work and set up the experiment, performed data collection and analysis under the supervision of author Dahl as part of a course in the Sound and Music Computing master program at Aalborg University. All authors participated equally in the writing of the final manuscript.

7. REFERENCES

[1] J. W. Davidson, "Visual perception of performance manner in the movements of solo musicians," *Psychology of music*, vol. 21, no. 2, pp. 103–113, 1993.

[2] M. Broughton and C. Stevens, "Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience," *Psychology of Music*, vol. 37, no. 2, pp. 137–153, 2008.

[3] S. Dahl and A. Friberg, "Visual perception of expressiveness in musicians' body movements," *Music Perception: An Interdisciplinary Journal*, vol. 24, no. 5, pp. 433–454, 2007.

[4] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin, "Cross-modal interactions in the perception of musical performance," *Cognition*, vol. 101, no. 1, pp. 80–113, 2006.

[5] W. F. Thompson and F. A. Russo, "Facing the music," *Psychological Science*, vol. 18, no. 9, pp. 756–757, 2007.

[6] W. F. Thompson, F. A. Russo, and S. R. Livingstone, "Facial expressions of singers influence perceived pitch relations," *Psychonomic bulletin & review*, vol. 17, no. 3, pp. 317–322, 2010.

[7] M. Schutz and S. Lipscomb, "Hearing gestures, seeing music: Vision influences perceived tone duration," *Perception*, vol. 36, no. 6, pp. 888–897, 2007.

[8] V. C. Tartter and D. Braun, "Hearing smiles and frowns in normal and whisper registers," *The Journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 2101–2107, 1994.

[9] J. J. Ohala, "The acoustic origin of the smile," *The Journal of the Acoustical Society of America*, vol. 68, no. 51, pp. 533–533, 1980.

[10] D. Huron, "Affect induction through musical sounds: an ethological perspective," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1664, p. 20140098, 2015.

[11] D. Huron, S. Dahl, and R. Johnson, "Facial Expression and Vocal Pitch Height: Evidence of an Intermodal Association," *Empirical Musicology Review*, vol. 4, no. 3, pp. 93–100, 2009.

[12] D. Huron and D. Shanahan, "Eyebrow movements and vocal pitch height: evidence consistent with an ethological signal." *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2947–52, 2013.

[13] Propellerhead Software AB, "Reason 9 Operation Manual," Tech. Rep., 2011.

[14] "Matlab Appdesigner," 2017. [Online]. Available: https://se.mathworks.com/products/matlab/app-designer.html

[15] A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd ed. Gainesville: John Wiley & Sons Inc., 2007, vol. 22, no. 1.

[16] D. J. Best and D. E. Roberts, "The Upper Tail Probabilities of Spearman's Rho," *Wiley for the Royal Statistical Society*, vol. 24, no. 3, pp. 377–379, 1975.

[17] J. Ohala, "Signaling with the Eyebrows–Commentary on Huron, Dahl, and Johnson," *Empirical Musicology Review*, vol. 4, no. 3, pp. 101–102, 2009.

[18] S. Chong, J. F. Werker, J. A. Russell, and J. M. Carroll, "Three facial expressions mothers direct to their infants," *Infant and Child Development*, vol. 12, no. 3, pp. 211–232, 2003.

[19] R. S. Kramer, J. E. King, and R. Ward, “Identifying personality from the static, nonexpressive face in humans and chimpanzees: evidence of a shared system for signaling personality,” *Evolution and Human Behavior*, vol. 32, no. 3, pp. 179–185, 2011.

[20] S. N. Geniole, A. E. Keyes, C. J. Mondloch, J. M. Carré, and C. M. McCormick, “Facing aggression: Cues differ for female versus male faces,” *PLOS one*, vol. 7, no. 1, p. e30366, 2012.