International Symposium on Performance Science
ISBN 978-94-90306-02-1

Facial expressions in vocal performance: Visual communication of emotion

Steven R. Livingstone[1], Caroline Palmer[1], Marcelo M. Wanderley[2], William Forde Thompson[3], and Jennifer Lissemore[1]

[1] Department of Psychology, McGill University, Canada
[2] Department of Music Research, Schulich School of Music, McGill University, Canada
[3] Department of Psychology, Macquarie University, Australia

This study investigated observers' emotional responses to facial expressions during song and speech. Silent presentations of facial expressions from neutral, happy, and sad song and speech productions were divided into three regions: prior to vocal onset, during vocal production, and following vocal offset. Observers were highly accurate and confident at identifying emotion during and following vocal production, but were less accurate for the region prior to vocal onset. Emotionally neutral presentations were identified less accurately and confidently than happy and sad presentations in all regions. Producers are known to exhibit decreased facial movement prior to vocal onset and for emotionally neutral utterances. These findings indicate that facial expressions may be important for the perception of emotion during vocal communication.

Keywords: facial expressions; emotion; perception; singing; speech

Facial expressions are integral components of non-verbal emotional communication and have been widely studied in a static posed context (Russell *et al.* 2003). However, normal conversation contains a variety of expressions that are rarely, if ever, static. Recent work (Livingstone *et al.* 2009) found that singers encode emotion using distinct facial movements that changed over the time-course of vocal production. Facial movements during vocalization differed from those prior to and following vocal production. It is unknown if these dynamic facial movements facilitate observers' perception of emotion.

We investigated observers' emotional responses to silent video recordings of facial expressions accompanying song and speech. Recordings were divided into three regions: prior to vocal onset (pre-production), during vocali-

zation (production), and following vocal offset (post-production). Two measures of response were analyzed: chosen emotion and response confidence. It was hypothesized that pre-production would be identified less accurately and less confidently, as this region typically contained significantly smaller facial movements (Livingstone *et al.* 2009). It was also hypothesized that production and post-production regions would be identified with similar accuracy and confidence, as both regions typically contained significant facial movements (Livingstone *et al.* 2009).

METHOD

Participants

Sixteen native English speaking adults (12 female), ranging in age from 18 to 37 years (mean=22.40, SD=4.47), were recruited from the Montreal area. Participants had received varied amounts of private music instruction (mean=4.81 years, range=0-12) and singing experience (mean=1.75 years, range=0-14). Two highly-trained female singers (model targets), with at least 9 years of vocal experience (first=10 years; second=9), were recruited from McGill University.

Materials and design

Singers were recorded (JVC Everio GZ-HD6 camera, AKG C 414 B-XLS microphone) while speaking or singing three neutral statements with one of three emotional intentions: neutral, very happy, or very sad. Statements were sung to a 0.3 ms inter-onset-interval (IOI) isochronous melody (F4, F4, G4, G4, E4, E4, F4). Recordings were divided into three vocal regions: pre-production (1.90 s prior to vocal-onset), production (vocal-onset to vocal-offset, mean duration=2.05 s; speech mean=1.62 s, song mean=2.48 s), and post-production (1.90 s after vocal-offset). Vocal regions were marked using Praat (Boersma and Weenink 2009), and recordings were edited using Adobe Premiere Elements. Video-only presentations (no audio) were presented to participants using E-Prime software. The within-subjects design contained 108 trials (2 vocalists x 2 production (speech/song) x 3 statements x 3 emotions x 3 regions).

Procedure and analyses

Participants were asked to rate the emotion of the performer using a 5-point bipolar scale (1=very sad, 2=sad, 3=neutral, 4=happy, 5=very happy), and their confidence of that judgment (1=very unsure, 2=unsure, 3=neutral,

4=sure, 5=very sure). Trials were blocked and counterbalanced by production, vocalist, and region (production region first), and randomized across statement and emotion. Emotion ratings were recoded for proportion correct responses (Sad=1, 2; Neutral=3; Happy=4, 5). Analyses were combined across vocalist and statement.

RESULTS

Viewers' mean accuracy is shown in Figure 1. The overall mean proportion correct scores was 0.93, which was significantly greater than chance values of 0.4 for happy and sad and 0.2 for neutral [$F_{1,15}$=7522.09, p<0.001]. A three-way repeated measures analysis of variance (ANOVA) revealed a significant main effect of emotion ($F_{2,30}$=9.45, p=0.001). Post-hoc comparisons (Tukey's HSD=0.07, α=0.05) confirmed that participants were less accurate for neutral (mean=0.87) than for happy (mean=0.98) and sad (mean=0.95). There was also a significant interaction of production with emotion ($F_{2,30}$=5.34, p=0.01) and of Production x Emotion x Region ($F_{4,60}$=3.09, p=0.022). Accuracy for speech-sad-pre-production (see Figure 1) appeared to be lower than other sad regions, suggesting a role in the 3-way interaction. Post-hoc comparisons (Tukey's HSD=0.13, α=0.05) confirmed that speech-sad-pre-production was less accurate than song-sad-pre-production, speech/song-sad-productions and speech/song-sad-post-productions.

As ratings for the production region were not originally hypothesized to be significantly different from post-production, a similar three-way repeated measures ANOVA was conducted, combining production and post-production regions. A significant main effect of region was reported ($F_{1,15}$=5.55, p=0.033), in which pre-production (mean=0.90) was significantly less accurate than prod-post-production (mean=0.94). There was also a main effect of emotion ($F_{2,30}$=6.93, p=0.003). A significant interaction of production with emotion was reported ($F_{2,30}$=7.52, p=0.002) as was emotion x region ($F_{2,30}$=4.28, p=0.023).

To confirm that ratings for production did not differ from post-production, a similar three-way repeated measures ANOVA was conducted, comparing only production with post-production (pre-production removed). No effect of region was reported. These results confirm that production and post-production regions had similar accuracies and were more accurate than pre-production.

Mean confidence ratings are shown in Figure 2. The overall mean confidence across conditions was 4.10/5 (82%). A three-way ANOVA revealed a significant main effect of emotion ($F_{2,30}$=14.05, p<0.001). Post-hoc compare-

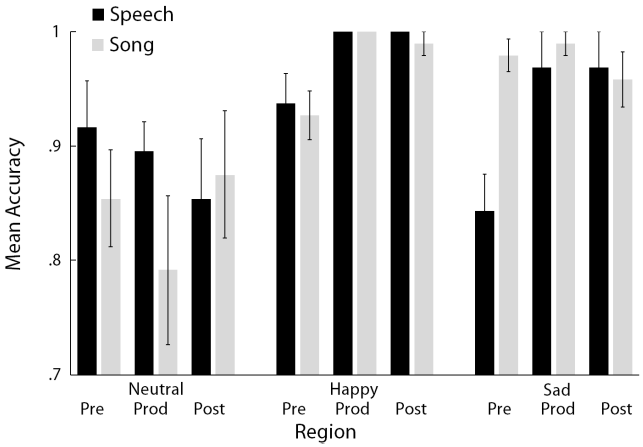

Figure 1. Mean proportion correct scores. Error bars denote the standard error of the means.

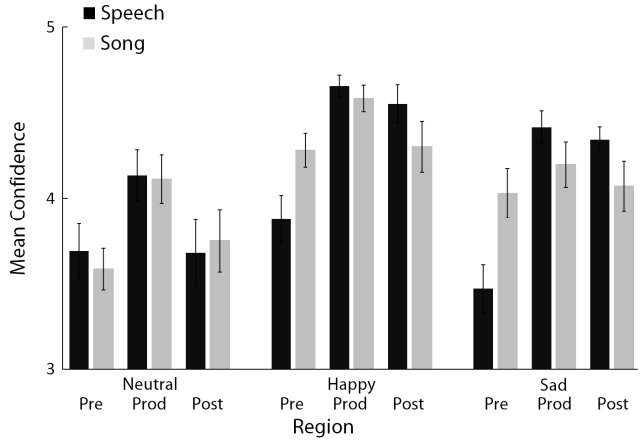

Figure 2. Mean confidence scores. Error bars denote the standard error of the means.

sons on emotion (Tukey's HSD=0.26, α=0.05) confirmed that participants were less confident for neutral (mean=3.83) than for sad (mean=4.09) and happy (mean=4.38), and less confident for sad than for happy. There was also a main effect of region ($F_{2,30}$=39.88, p<0.001); post-hoc comparisons on region (Tukey's HSD=0.18, α=0.05) confirmed that participants were less confident for pre-production (mean=3.82) than for post-production (mean=4.12) and production (mean=4.35), and less confident for post-production than for

production. There was also a significant interaction of production with region ($F_{2,30}$=14.00, p<0.001) and production x emotion x region ($F_{4,60}$=5.31, p=0.001).

Confidence for speech-sad-pre-production (see Figure 2) appeared to be lower than other sad regions, suggesting a role in the 3-way interaction. Post-hoc comparisons (Tukey's HSD=0.46, α=0.05) confirmed that speech-sad-pre-production was less confident than song-sad-pre-production, speech/song-sad-productions, and speech/song-sad-post-productions. Confidence for speech-happy-pre-production also appeared to be lower than other happy regions. Post-hoc comparisons confirmed that speech-happy-pre-production was less confident than speech/song-happy-productions and speech-happy-post-production. Similarly, song-neutral-pre-production was less confident than speech/song-neutral-productions. These results suggest that participants were consistently less confident in their evaluation of pre-production than production and post-production regions.

Participants' accuracy and confidence results suggested a relationship between the measures. A Spearman's rank correlation between mean accuracy and confidence responses was significant (r=0.80, p<0.01), indicating that when participants were more accurate they were also more confident.

DISCUSSION

Observers' emotional responses to facial expressions during speech and song were highly accurate. These results were achieved in the absence of auditory information, suggesting that facial expressions may play an important role in the perception of emotion before, during, and after vocal communication.

Observers were also accurate and confident at identifying emotions from facial expressions that occurred after the offset of vocal production. This suggests that speakers and singers continue to sustain facial expressions long after vocal production has ended and that this behavior accurately conveys emotional information (Livingstone *et al.* 2009). Observers were significantly less accurate and less confident at identifying emotions prior to the onset of vocal production, supporting the initial hypothesis. This hypothesis stemmed from previous work which reported that pre-production exhibited significantly smaller facial movements than production and post-production (Livingstone *et al.* 2009). Emotionally neutral presentations were identified less accurately and less confidently than either happy or sad recordings. It is thought that the decreased facial movements exhibited by neutral vocal productions (Livingstone *et al.* 2009) elicited these results. One potential weak-

ness of the study was that accuracy ratings were all close to ceiling for the facial expressions used in the study. Future work will address this concern.

This research demonstrates that facial expressions that accompany speech and song may facilitate the perception of emotion and, in the absence of sound, are sufficient for robust emotional communication.

Acknowledgments

This work was supported by an NSERC-Create fellowship to the first author, the Canada Research Chairs program and NSERC Grant 298173 to the second author, and an ARC discovery grant DP0987182 to the fourth and second authors. We thank Frances Spidle, Pascal Lidji, Rachel Brown, Michele Morningstar, and Max Anderson for comments.

Address for correspondence

Steven Livingstone or Caroline Palmer, Department of Psychology, McGill University, 1205 Dr Penfield Ave, Montreal H3A 1B1, Canada; *Email:* steven.livingstone@mcgill.ca or caroline.palmer@mcgill.ca

References

Boersma P. and Weenink D. (2009). Praat: Doing phonetics by computer. Version 5.1.05 ed.

Livingstone S. R., Thompson W. F., and Russo F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception, 26,* pp. 475-488.

Russell J. A., Bachorowski J.-A., and Fernández-Dols J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* pp. 329-350.