The Influence of Facial Expressions and Vocal Expressions on The Perceived Expressiveness.

T. Anthonissen, S. Brands, L. van Dinteren, H. van der Kallen, J. Moonen, A. Pieterman

Abstract

Emotions can be expressed in different levels of expressiveness, and can be recognized based on facial and vocal cues. A perception test was performed in order to examine whether the perceived level of emotional expressiveness depends most on facial expressions, vocal expressions or both. Thirty fragments of Jimmy Kimmel's Celebrities Read Mean Tweets were shown in three conditions, namely: audio only, video only and audio and video. All fragments were rated on the perceived level in which the emotions anger, contempt, disgust, fear, happiness, sadness and surprise were expressed. The findings suggest that there is no difference between facial and vocal expressions in the perception of emotional expressiveness. The data was also analyzed by the Computer Expression Recognition Toolbox (CERT), which showed that the emotion *contempt* was the most prominent present emotion in the stimuli. This research suggests that expressiveness is difficult to measure over multiple emotions, but for a single emotion it was shown that facial expressions were perceived more expressive than vocal expressions.

Keywords: perceiving emotions; expressiveness; non-verbal; facial expressions; vocal expressions.

Introduction

Facial and vocal expressions are essential aspects in human interaction (Russell, Bachorowski & Fernández-Dols, 2003). The receivers' ability to analyze these expressions provide the opportunity to recognize someone's emotion. The face is seen as the window with a view opening onto our emotions (Russell & Fernández-Dols, 1997). Ever since the 1980s psychologists said that the face is the key to understanding emotion, and emotion is the key to understanding the face. That is why the link between facial expressions and emotional experience are widely investigated by researchers (Tottenham, et al., 2009). However, vocal expressions are rich with cues which convey the emotional state. In speech, speakers provide their emotional state through acoustic properties (Bachorowski, 1999). Each emotion is associated with a distinctive tone of voice (Bachorowski, 2008). Even in absence of facial cues, the receiver is able to make an accurate evaluation about the emotional state because of the distinctive tone of voice (Bachorowski, 1999). With evidence that both facial and vocal expressions provide clarity about someone's emotion, the question arises: is one of the expressions more dominant than the other?

For both facial expressions (Elfenbein & Ambad, as cited in Russell, Bachorowski, & Fernández-Dols, 2003) and vocal expressions (Johnstone & Scherer as cited in Russell, Bachorowski, & Fernández-Dols, 2003), evidence is found that despite differences in culture, age or background, receivers often agree on emotion signals. However, previous studies focused mostly on either facial or vocal expressions rather than a combination (Russell, Bachorowski & Fernández-Dols, 2003). There is a study of Wallbott and Scherer (as cited in Russell, Bachorowski, & Fernández-Dols, 2003) that states that the agreement on emotion signals is higher for facial than for vocal expressions. Even though this does not prove that facial expressions are more dominant than vocal expressions, it might be an indicator.

The theory of Tomkins is likewise based on both vocal and facial expressions (Russell, Bachorowski, & Fernández-Dols, 2003). According to this theory, each basic emotion can vary in intensity and consists of a single brain process. Triggering an emotion will trigger all facial and vocal expressions belonging to this emotion. However, the question remains if facial and vocal expressions always correlate on the emotion that is recognized. Is it possible that the perceived expressiveness of emotions in facial expression differs from those in vocal expression?

This research will investigate both vocal and facial expressions. The research question that will be investigated is:

RQ: *What is the influence of facial and vocal expressions on the perceived expressiveness?*

The following hypothesis is designed following the described studies:

H1: *Facial expressions are perceived more expressive than vocal expressions.*

Stimuli Collection

Selection Criteria and Procedure

In the current study, the expressiveness is measured through an online perception test. The chosen fragments were derived from Jimmy Kimmel's Celebrities Read Mean Tweets, which consist of spontaneous reactions of celebrities reading mean tweets about themselves. In total 30 fragments of American celebrities were selected. These fragments had to meet certain conditions. Based on the duration of the fragment, short fragments (<7 seconds) were not used for the analysis. The celebrities' face had to be fully visible and the chosen fragments consisted of both male and female celebrities, so the results could be generalized for both sexes.

The used fragments were chosen for several reasons. Firstly, the celebrities read these tweets for the first time, so it is assumed they behave naturally. Furthermore, every fragment consists of the same 'goal', namely reading a mean tweet about themselves, which makes the fragments unambiguous. Since the celebrities' faces were fully shown, it was possible to measure the facial expressions with the Computer Expression Recognition Toolbox (CERT). Finally, the fragments are meant as entertaining fragments. This might lead to a higher attention span of the participants who have to judge the presence and strength of several emotions. These arguments make the fragments suitable for measuring expressiveness.

Video Editing

The episodes were cropped into fragments of one celebrity to ensure that the focus is on that particular celebrity. Cropping the episodes into fragments was done with Windows Movie Maker. The average duration of the fragments is ten seconds (minimum 7 seconds and maximum 15 seconds). The audio-video condition consisted of both video and audio. In the video only condition, the audio information was removed. In the audio only condition all visual information was removed, except the tweet. The tweet was added on the screen to ensure that what was said in the audio information was clear for the participants to judge the expressiveness. The links to the fragments of all conditions are included in Appendix I.

Online Perception Test

Participants

This study consisted of 60 Dutch persons in which 26 males and 34 females participated. The average age was 25.3 years (SD = 7.2) with a minimum of 19 years and a maximum of 57 years. All participants were randomly selected and assigned to one of the three conditions. Every participant was sufficient in the English language, which is an important aspect since English was the language that was used in all stimuli.

Procedure

In this study, 30 fragments of the Celebrity Read Mean Tweets were shown in all of the three conditions (audio, video and audio-video) and judged by 60 participants. The participants were equally divided over each condition, which means that the fragments in every condition were judged by twenty participants. The fragments and accompanying questionnaire were presented in Qualtrics and sent to the participants digitally. The participants performed the tasks while using their own computers at home and were instructed to locate themselves in a silent environment, so they would not be disturbed by others.

With regard to the task, the participants were instructed to listen (in the audio condition), watch (in the video condition) or listen and watch (in de audio-video condition) to the fragments and to judge the extent in which the emotions shown by the celebrities were. Before the main task began, the participants were asked to answer demographical questions about their gender, age and educational level. In addition, questions were asked about their sufficiency in the English language and their acquaintance with the Celebrities Read Mean Tweets from the Jimmy Kimmel show. After these questions were answered, the task began. The task consisted of viewing and/or listening to the fragments and answering questions about the perceived extent in which emotions were expressed by the celebrities. All questions were asked in Dutch. The emotions were: anger (boosheid), contempt (minachting), disgust (walging), fear (angst), happiness (blijdschap), sadness (verdriet) and surprise (verrassing) (Matsumoto, LeRoux, & Wilson-Cohn, 2000). The participants rated the perceived expressiveness of these emotions on a 7-point Likert scale (1 = the emotion is totally not present, 7 = the emotion is very much present). The same questions were answered after every fragment appeared. The order in which the fragments were presented was randomized. There was no practice material.

Results Online Perception Test

Analysis 1

The expressiveness of the reviewer was measured with seven items on a highly reliable 7-point scale (α = .89). For the first analysis, three new variables were computed, since the items turned out to form a reliable scale. These three variables were the means of the perceived expressiveness for the audio-video fragments, audio only fragments, and video only fragments. The data did not contain any outliers and therefore no data were excluded from the analysis. The analysis of normality showed that the three new calculated variables were normally distributed (the largest z-score was 1.07). There were no significant differences in the variances of the three groups, as indicated by the Levene's test (F(2, 57) = 0.93, p = .402), so the assumption of homogeneity of variance is met.

On average, participants rated celebrities in the audio-video condition with an expressiveness of 2.69 (SD = 0.66), celebrities in the audio only condition with an expressiveness of 2.64 (SD = 0.83), and celebrities in the video only condition where on average rated 2.71 (SD = 0.76) on expressiveness. To test the hypothesis that facial expressions are more dominant in expressiveness than vocal expressions, an ANOVA was conducted. Unfortunately, the ANOVA showed no significant differences between the three conditions (F(2, 57) = 0.05, p = .948). This means that no support has been found for a difference in the degree of expressiveness between the three conditions.

Analysis 2

To investigate if there were differences between the fragments, all 30 fragments were analyzed separately. Thirty new variables were calculated, with the mean of expressiveness per celebrity. A dummy variable was added

to split the variables in the audio, video and audio-video condition. Two variables had some slight kurtosis in the audio-video condition (z-scores 2.15 and 2.04). Since this is such a small value, it is only a small part of the data, and all other data were normally distributed, an ANOVA was conducted. There was one variable (David Cross) with significant differences in the variances, as indicated by the Leven's tests of all thirty variables. Therefore, for this one variable the significance score was indicated by the outcome of Welch's test. The ANOVA showed no significant differences between the three conditions in all thirty variables with the lowest p-value of $p = .06$. This means that no support has been found for a difference in the degree of expressiveness per fragment between the three conditions.

Analysis 3

In addition, it might be interesting to investigate if there was a correlation between the three conditions (audio-video, audio, and video). In this analysis three variables were used; the three variables that were calculated for analysis 1, with the means of expressiveness in all three conditions. As can been seen in the results of analysis 1, the data were normally distributed. Therefore, a Pearson correlation was conducted to investigate a relationship between the three variables. The Pearson correlation showed a (slight) significant correlation between the three conditions (audio and video ($r(30) = .392$, $p = .032$), audio and audio-video ($r(30) = .416$, $p = .022$), and video and audio-video ($r(30) = .368$, $p = .045$)). These results indicate that the three conditions were mutually reinforcing. The video only condition in combination with the audio-video condition was judged more expressive than the audio only condition in combination with the audio-video condition. Since there was only a slight significance, it might be interesting to investigate if there are larger significant differences in each emotion separately.

Analysis 4

With a MANOVA was analyzed which emotion was most present. Unfortunately, no significant difference was found among the seven emotions. This means that no support has been found for one emotion being more present than other emotions.

Video Analysis

Procedure

In analysis 4 no significant result was found. Therefore, CERT was used to analyze which emotion was most present in the fragments, based on facial expressions. Only the emotion that appears to be most present was further analyzed. Investigating one emotion separately broadens and deepens the present study.

Analysis 5

In order to determine the emotion that was most present in the fragments, CERT was used. Table 1 shows that the emotion *contempt* has the highest mean for both intensity ($M = 0.32$, $SD = 0.10$) and frequency ($M = 0.07$, $SD = 0.05$) compared to the other emotions.

Table 1
*Statistics for the emotions analyzed by CERT (*N=*30).*

	Intensity		Frequency	
Analyzed emotion	Mean	Standard deviation	Mean	Standard deviation
Anger	0.25	0.20	0.04	0.05
Contempt	0.32	0.10	0.07	0.05
Disgust	0.16	0.23	0.02	0.04
Fear	0.09	0.17	0.01	0.02
Joy	0.24	0.26	0.06	0.09
Sad	0.28	0.15	0.06	0.06
Surprise	0.16	0.21	0.02	0.05
Neutral	0.30	0.09	0.06	0.06

Analysis 6

The correlation was measured between the three conditions for the emotion *contempt*. Table 2 shows the correlation between the conditions. All conditions correlate significantly. There is a positive correlation between audio-video and video ($r = 0.781$, $p < .001$), a positive correlation between audio-video and audio ($r = 0.658$, $p < .001$) and a positive correlation between audio and video ($r = 0.384$, $p = .033$).

Table 2
*Descriptive statistics and correlations for the predictors and outcome (*N = *30).*

	Mean	SD	Audio +Video	Audio	Video
Audio +Video	3.26	0.90	*1.00*		
Audio	3.52	0.59	*.66***	*1.00*	
Video	3.22	0.61	*.78***	*.38**	*1.00*

** Correlation is significant at the .01 level (2-tailed).
* Correlation is significant at the .05 level (2-tailed).

Discussion

In this research, the main goal was to assess whether facial expressions are perceived as more expressive than vocal expressions. In order to measure these differences, participants were either divided in the audio only, video only, or audio and video condition in which YouTube fragments of the Celebrity Read Mean Tweets were shown.

The overall results showed that the hypothesis is not supported, since no differences were found between the three conditions (audio, video, and audio and video). These results contradict previous findings regarding the emotional signals in facial and vocal expressions. It is suggested by several researchers that the emotional signals are higher for facial than for vocal expressions (Hess et al., 1988; Wallbott & Scherer, 1986). A possible explanation is that in all the conditions the tweet which was read by the celebrity was visible. Other results might occur when the tweet is not provided in all of the conditions, since participants might have focused on the texts rather than evaluating the expressiveness of the celebrities.

Although the results did not indicate a difference for overall expressiveness, a difference was found between the three conditions and the perceived emotion *contempt*. This type of emotion was evaluated, since the results of CERT indicated that this was the most expressed emotion of the celebrities. The results indicated that the expressiveness of the emotion *contempt* was more perceived in video than in audio, which partially supported the hypothesis. These results do agree with previous findings of Hess et al. (1988) and Wallbott and Scherer (1986) that the emotional signal is higher for facial than for vocal expressions.

A limitation of this research is that tweets were read by celebrities, mainly actors and musicians. This means that they are experienced in acting in front of a camera and they may have been trained in expressing emotions. Furthermore, the celebrities were not speaking spontaneously, since they were reading the tweets. This might have influenced their reactions. Since the celebrities were informed about the mean content of the tweets, they might put on an 'act' and provide less facial and vocal expressions. Therefore, it is suggested that future research should take this into account by using no celebrities, but, for example, students or other participants who have no experience performing in front of a camera. In addition, it is suggested to not solely evaluate the emotion *contempt*, also the other emotions could be researched in more detail.

Conclusion

The aim of the study was to examine what the influence of facial and vocal expressions was on the perceived expressiveness. The findings showed no significant differences between the audio only, video only, and audio and video condition concerning people's perceived expressiveness. However, a significant difference was found for one type of emotion, namely *contempt*. This result indicated that people perceive the expressiveness of the emotion *contempt* more in the video condition than in the audio condition.

References

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current directions in psychological science*, *8*(2), 53-57.

Bachorowski, J. A., & Owren, M. J. (2008). Vocal expressions of emotion. *Handbook of emotions*, *3*, 196-210.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods*, *42*(1), 351-362.

Matsumoto, D., LeRoux, J., & Wilson-Cohn, C. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, *24*, 179 –209

Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual review of psychology*, *54*(1), 329-349.

Russell, J. A., & Fernández-Dols, J. M. (1997). 1. What does a facial expression mean?. *The psychology of facial expression*, 1.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ... & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research*, *168*(3), 242-249.

Appendices

Appendix I links to fragments

1. Andy Dick
   Video en Audio: https://www.youtube.com/watch?v=9sMSHXDK6cs
   Video: http://youtu.be/bcSX0B9OIpI
   Audio: http://youtu.be/dWhmBxwTihI
2. Anne Faris
   Video en Audio: https://www.youtube.com/watch?v=gJi1647Qqhk
   Video: http://youtu.be/5HiOMeKfxR0
   Audio: http://youtu.be/AS5NA6RJguE
3. Busy Phillips
   Video en Audio: https://www.youtube.com/watch?v=LIFDvfp7ZTA
   Video: http://youtu.be/cevvPJEXzSs
   Audio: http://youtu.be/74GY1k5tvKM
4. David Cross
   Video en Audio: https://www.youtube.com/watch?v=hrX-W1s4h94
   Video: http://youtu.be/SeYxtGHX8YY
   Audio: http://youtu.be/976R2xRyDzc
5. Jason Bateman
   Video en Audio: https://www.youtube.com/watch?v=nrxV_l5Kkpw
   Video: http://youtu.be/SWA8-kfg1h8
   Audio: http://youtu.be/tixWUPdtCuA
6. Joel Mchale
   Video en Audio: https://www.youtube.com/watch?v=AU-gYn9QL6k
   Video: http://youtu.be/lNaSdYTDk5w
   Audio: http://youtu.be/E1hYEH2G8mI
7. James van der Beek
   Video en Audio: https://www.youtube.com/watch?v=M70kh8yl6Rs
   Video: http://youtu.be/MYQ-vlbihNI
   Audio: http://youtu.be/whXJW4iigrc
8. Larry King
   Video en Audio: https://www.youtube.com/watch?v=w37Pl4Iv73I
   Video: http://youtu.be/uT_Zu1M0_TQ
   Audio: http://youtu.be/6_jCcBdVtGY
9. Jessica Biel
   Video en Audio: https://www.youtube.com/watch?v=v6bNZTJpV2U
   Video: http://youtu.be/2y9X-GyziyE

Audio: http://youtu.be/IF9CS2awPDM

10. Werner Heisenberg
    Video en Audio: https://www.youtube.com/watch?v=bnHCkcj3Mxw
    Video: http://youtu.be/pO5NyxVVQw4
    Audio: http://youtu.be/p3YNv30M49Q
11. Andy Samberg
    Video en Audio: https://www.youtube.com/watch?v=A7VG5yyge9w
    Video: http://youtu.be/sH_Yt_KKMTc
    Audio: http://youtu.be/cUU8j9hEnWo
12. Jessica Alba
    Video en Audio: https://www.youtube.com/watch?v=A8wuOz8rO0o
    Video: http://youtu.be/LZ5XxmtLPyk
    Audio: http://youtu.be/u8mbNTp_0zo
13. Dennis Quaid
    Video en Audio: https://www.youtube.com/watch?v=AKZJ6cYGKLs
    Video: http://youtu.be/N_SLQ9QSQ4s
    Audio: http://youtu.be/XBD-Z8hklk0
14. John Krasinski
    Video en Audio: https://www.youtube.com/watch?v=UT7gaSveM7U
    Video: http://youtu.be/a7St9S93JY8
    Audio: http://youtu.be/RbN1YXykNl0
15. Mark Ruffalo
    Video en Audio: https://www.youtube.com/watch?v=N7zETBThPuQ
    Video: http://youtu.be/qxeLlFVZJHM
    Audio: http://youtu.be/rI8RXAOI0OI
16. Sharon Stone
    Video en Audio: https://www.youtube.com/watch?v=uogP8GXRFhM
    Video: http://youtu.be/W2jhL-tB8Ac
    Audio: http://youtu.be/WSpik2do2Os
17. John Goodman
    Video en Audio: https://www.youtube.com/watch?v=9J0h_6McngU
    Video: http://youtu.be/t796Yz2LSHc
    Audio: http://youtu.be/2ZNGSwFFVXE
18. John Hamm
    Video en Audio: https://www.youtube.com/watch?v=OS2LrhErnTg
    Video: http://youtu.be/67wRQ6B__z4
    Audio: http://youtu.be/5H7xyFlUgxg
19. Sarah Silverman
    Video en Audio: https://www.youtube.com/watch?v=G_tfL5j9QGw
    Video: http://youtu.be/NNuZQPpXKt0
    Audio: http://youtu.be/BbSqX0KBMD8
20. Ethan Hawke
    Video en Audio: https://www.youtube.com/watch?v=NYuXUozAFXU
    Video: http://youtu.be/NAYGru0Wtzw
    Audio: http://youtu.be/LWs_FnWbR5c
21. Julia Roberts
    Video en Audio: https://www.youtube.com/watch?v=mejG6vqsASs
    Video: http://youtu.be/D7VHC3Q5-Uo
    Audio: http://youtu.be/IlyCx9a-xJ8
22. Kit Harrington
    Video en Audio: https://www.youtube.com/watch?v=Y0SfOE16Drg
    Video: http://youtu.be/i83EczPGwwU
    Audio: http://youtu.be/pG7w06hpo4M
23. Mindy Kaling
    Video en Audio: https://www.youtube.com/watch?v=MgtJQqXxRJI
    Video: http://youtu.be/UVbxuLi90sk
    Audio: http://youtu.be/pcBAybcMmHU
24. Demi Lovoto
    Video en Audio: https://www.youtube.com/watch?v=QOiJdp8Qauo
    Video: http://youtu.be/ekFe40MXGlo
    Audio: http://youtu.be/JPIP3HFOpMg
25. Ariana Grande
    Video en Audio: https://www.youtube.com/watch?v=_abol9LsCtw
    Video: http://youtu.be/1I0woL-P-HE
    Audio: http://youtu.be/AZxI0ItNix4
26. Britney Spears
    Video en Audio: https://www.youtube.com/watch?v=luvfqXsf7Uw
    Video: http://youtu.be/J9ZOVHcx4Xc
    Audio: http://youtu.be/IrReHmz8guU
27. Drake
    Video en Audio: https://www.youtube.com/watch?v=J7MQlYv809w
    Video: http://youtu.be/lqkLghwK1k8
    Audio: http://youtu.be/fqM4z4KrAY0
28. Josh Groban
    Video en Audio: https://www.youtube.com/watch?v=2gdR2zU4qSk
    Video: http://youtu.be/P3vGZdhpsJE
    Audio: http://youtu.be/zTuwy9WERKA
29. Katy Perry
    Video en Audio: https://www.youtube.com/watch?v=snm4GV_tM_k
    Video: http://youtu.be/V3jBBdALB_c
    Audio: http://youtu.be/JLp0cZAVyHs
30. Sam Smith
    Video en Audio: https://www.youtube.com/watch?v=ijpXAruDVzg
    Video: http://youtu.be/E_iRsWvLwe8
    Audio: http://youtu.be/MAujIE3BOZY