

The facial and vocal expression in singers: a cognitive feedback study for improving emotional expression in solo vocal music performance

António Salgado (Universidade de Aveiro, Portugal)

ABSTRACT: The paradigm used in most of the studies exploring how singer's express emotional states has been, since Seashore's (1947) pioneer work and Kotlyar & Morozov's (1976) study, basically the same. Singers are asked to sing a melody or a number of short melodies expressing different emotions according to the researcher's instruction. The performed melodies are firstly recorded or videotaped and then listened to, watched and evaluated in order to check whether listeners are able to recognise the intended expression. Each performance may then be further analysed in order to study what acoustic or visual means the singer has used to achieve each emotional expression, but this, to date, has only very recently been explored (Salgado, 2001, 2003).

The investigation here undertook aimed to study the perception and recognition of emotional meaning within the performance of singing in a real concert situation. That is, how a certain audience perceives and recognises the emotional meaning communicated through the voice and face of the singer within the real-time life-performance of singing.

The performer learned and sang the Lied 'Die Post' from Schubert's cycle 'Die Winterreise'. Therefore, the Lied was not only previously analysed in itself, but as an integrated part of the general context of the proposed song cycle. The intended interpretation of the Lied was then communicated to the accompanying pianist. The music moments where the emotional peaks were supposed to happen were linked to the places where in the musical structure the pianist should underline the correspondent musical features. Fifteen participants were asked to observe the performances of the Lied and to judge the musically conveyed emotional meanings. They were seated by a computer screen, and they were asked to rate in real-time feedback the emotional content, or the different emotional contents, occurring during the proposed performed interpretation by pressing emotion buttons indicated on the computer keyboard. The choice of the descriptive terms was also in this experiment limited to the four main emotions found by the musical-poetic analysis of the referred song. Each group of five screens, for technical reasons, was connected to a distribution post (Posto 1, Posto 2, and Posto 3). To each distribution post was attributed one colour: red, blue and green. So, each group of five observers' answers appeared in the graph laid out red, or blue, or green coloured. This song was performed twice, and on each occasion the audience was asked to rate the emotional meaning over the course of the performance. Each listener pressed the key according to the emotion perceived previously chosen (S-sadness; H-happiness; A-anger; F-fear). The answers were given in a real-time during the performances, and the emotion button was held down as long as the Emotion was perceived.

From the first to the second run through participants rated significantly fewer of the 30 moments as neutral and rated significantly more as representing fear. Fear is identified for significantly more time during the 4 moments in run through two than in run through one. Overall participants make significantly more correct key presses in run through two than one during these peak periods and the errors are accounted for mainly by neutral key presses - that is there were significantly more neutral presses in run through 1 than 2 during the peak emotional periods and the total correct plus neutral errors accounts for the 16 possible moments. While there was a suggestion that in run through one fear was more easily confused with anger than in run through 2 this was not confused for significantly more key moments for run through 1 than 2.

According to the graphs summarising the final results, it can be seen that there was a clear improvement in the perception and recognition of the emotional meaning perceived in all the emotions expressed and especially the clarification of the expression of the emotion of fear. So, the conclusion might be that it is with the rehearsal and study of the different videotaped performances and the results of the feedback from the audience's responses for each performed version of the same song, that have shown which facial and vocal cues might be used in order for a performer be able to explore his/her

Background

The two main approaches to understand and classify emotional behaviour are the dimensional approach and the categorical approach. The dimensional approach focuses on the identification of emotions based on their placement on a two, or three, dimensional structure (valence, activity, and potency). Russell's (1980) dimensional approach is a circumplex model of emotion - a two dimensional, circular structure, based on the bi-polar dimensions of valence (like Vs dislike) and arousal (high Vs low). Analysis of judgements of emotion words or facial expressions using factor analysis or multidimensional scaling is the most common of this kind of approach. The purpose of Russell's (1980) model and other models or formulations (Wundt, 1987; Woodworth, 1938; Schlosberg, 1941) seem, in fact, more oriented to investigate music mood-induction than perceived emotional categories. According to the categorical approach, listeners experience emotions as categories that are distinct from each other. The essential aspect for the defenders of the categorical approach (see Tomkins, 1962; Izard, 1977; Ekman, 1992; Oatley, 1992; Plutchnik, 1994) is the concept that basic emotions are of a limited number, innate and universal, and all the other emotional states derive from them. According to Scherer (1995:235), "the most important feature of the emotion mechanism is that it produces specification readiness while providing a latency period that allows adaptation of the behavioural reactions to the situational demands." So, they can be seen as adaptive in life emergency situations (ready for execution) but also as way of externalising a reaction to a specific emergent situation, and a way of communicating this information to the social environment. Frick (1985) and Scherer (1995) claimed, based on listener's ability to correctly recognise vocal emotional expressions, that the vocal expression of at least some emotions (a list where all the basic emotions are included but also love, pride, jealousy, etc) seems to be universal. Scherer (1981) presented a review of different studies that appeared over the last 50 years on the recognition of emotional states from voice samples and reported an average accuracy of ~ 60%. Bezooijen (1984), for instance, reported a vocal emotion recognition mean accuracy of 65% for emotions such as, disgust, surprise, shame, interest, joy, fear, sadness and anger. Later, Scherer et al. (1995) reported for fear, joy, sadness, anger and disgust, a mean accuracy of 56%. Salgado (2001) reported a mean accuracy for basic emotions' vocal expression in singing of ~ 65%. The paradigm used in most of the studies exploring how singer's express emotional states has been, since Seashore's (1947) pioneer work and Kotlyar & Morozov's (1976) study, basically the same. Singers are asked to sing a melody or a number of short melodies expressing different emotions according to the researcher's instruction. The performed melodies are firstly recorded or videotaped and then listened to, watched and evaluated in order to check whether listeners are able to recognise the intended expression. Each performance may then be further analysed in order to study what acoustic or visual means the singer has used to achieve each emotional expression, but this, to date, has not been explored. Hence, it is an aim of the current investigation. The use of real music excerpts to study emotional expression in singing ensures good ecological validity (see Gabrielsson and Lindström, 2001). Also, the common singing of the same melody with different kinds of emotional expression secures the internal validity of the experiment, because the recognition of the conveyed emotion should be mainly the result of the singer's expressive intentions (a device used by Davidson (1993) in the exploration of expressive intention). All these experiments have focused, for the reasons exposed, on the perceived emotion rather than on the induced emotion. The emotions under investigation have usually been sadness, happiness, fear, anger, and tenderness. The studies have mainly focus on two things: the accuracy of the communication and the code used by the singers and the listeners (see Juslin, 2001) to represent the emotion. Of course, the different ways people express, recognise and experience emotions is certainly of capital importance in the manner how subjects differentiate emotions and emotional contents within artistic communication. But, given the assumptions that the face

is 'the primary theatre' of emotions (Ekman, 1992), and that either dimensional and categorical approaches take facial expressions as a powerful source of emotional data, and that the performance of Lied (song) is basically expressive at the level of the face (facial expressions, movements and articulation), it would be expected to find much and interesting evidence for complementary work of these different investigation approaches applied to music performance. Nevertheless, the main guidance of this investigation was based on the possible contribution of the categorical approach alone, once the methodology followed and used the experiment was based on multiple-choice questionnaires and verbal and non-verbal recognition of basic emotions chosen from word-concepts or prototypical facial expressions taken from Ekman and Friesen's (1976) still photographs of basic emotions.

Aims

The investigation here undertaken aimed to study the perception and recognition of emotional meaning within the performance of singing in a real concert situation. That is, how a certain audience perceives and recognises the emotional meaning communicated through the voice and face of the singer within the real-time life-performance of singing. The aim of this study was also to investigate how to elaborate an expressive tool, based on a model of a real-time life-performance cognitive feedback, that might help singers and singing teachers to test and improve the expressive capacity of themselves and of the students of singing. It seems important that in the performance of music in which structure clearly points to an interpretation of categorical emotional meaning, the singer might be able to convey that meaning accurately by using facial and vocal cues that might contribute expressively to the communication of the intended emotional content. So, it seemed relevant to elaborate a 'tool' that can give immediate cognitive feedback of the degree of expressiveness of a real-time performance, i.e., an "expressive tool" where the singer's ability to express emotional meaning could be tested and improved. This has been simultaneously the aim and the process of the experimental investigation undertaken: the elaboration and testing of a reliable cognitive feedback for a real-time life-performance that could function as an "expressive tool" for the use of the singers.

The hypothesis underpinning the investigation was: if, during the course of this study, the perceivers' recognition of the singer's emotional interpretation could enable a clearer perception of expressiveness and, consequently, to help to improve the expressiveness of performance, then it might be possible to conclude that the process used in this experiment would have accomplished its aim and that it might aid other singers and singing students to improve the way they use their voices and their faces when interpreting emotionally the musical structure.

So, before the main experimental procedure was undertaken four steps were undertaken:

1. The musical structure of the referred song and the poetic content of the words were analysed and integrated as a part of the general context of the proposed song cycle.
2. A performance was worked out so that it could be able to corroborate the emotional content found in the musical-poetic analysis.
3. Computer software was designed to create a reliable way to register the audience feedback to the performer's interpretation.
4. An experimental context was developed elaborated, where the planned performance, like in a real time concert situation, was assessed through an audience's real-time feedback. Therefore, and according to Schubert (Schubert, 2001:409), this experiment might be considered as an ecologically valid experiment, that is the listeners were listening to an

uncontrolled, real piece of music rather than an excerpt or contrived auditory stimulus.

The development of the experimental-study

As mentioned, the first issue to be considered was a software design to programme a way to reliably register the audience's categorical recognition of the performed emotions. The methodology devised aimed to create a situation of an almost continuous experiment. Of course, as Schubert (2001:395) explains, "categorical responses require the participant to select one or several answers from a range of answers. (...) In a continuous response task, this may either interfere with the listening experience, or lead to the omission of some of the responses or guessing so as to minimise interference." Nevertheless, the target of this experiment was not to investigate all the emotional continuous nuances that the musical structure of the song might eventually convey. In fact, the experiment aim was to investigate if the performer, according to the analysed structural meanings and within an ecologically valid framework, was able to communicate accurately, during the continuous development of the musical structure of the song, all the previously selected emotional contents. Therefore, categorical responses were made, but over time, and according to intentional peaks of emotional expression performed, which have been programmed and rehearsed according to the analysis of the musical structure and the interpretation of the song.

In this experiment, the performer aimed to convey four different emotional peaks, which corresponded to the piece's structure and the four emotional contents (happiness, fear, sadness, and anger). So, in this experimental study what has been judged as 'emotional stimuli' was not the musical structure itself, but the singer's performance of the song, i.e., the singer's capacity to convey vocally and facially the four most relevant emotional contents analysed in the song. The main questions to be answered in this experiment were:

- How, or by which means, has the performer decided to convey the chosen emotional meaning?
- Which, musical elements and features were used by the performer to convey the emotional?
- To what extent has the audience been able to rate the performer's expressiveness?
- And, how (or if) the performer has been able through the audience's feedback to improve the expressiveness of his performance?

Therefore, the aim of this experimental study was, as said, to validate and improve the performer's communication through the audience's identification and recognition of the emotional musical contents portrayed during the performance.

Material used

For the development of the experimental procedure the software and hardware used were:

Software:

Macromedia Director 8.0
Xtra - FILE I/O

HardWare:

Multimedia Personal Computer
iMac (for the elaboration of the Macintosh version)

For the experimental procedure itself, the material used consisted of:

Infrastructure:

ETHERNET (10 Mbits) with limited access to the 4 computers involved in the experiment.
Hub 8 doors

Server:

Multimedia Personal Computer.

(responsible for generating the logs in real time and for the laying out of graphs with the audience's answers)

3 distribution posts:

Multimedia Personal Computer IBM

(each one constituted of)

5 Monitors Samsung

5 Stereo headphones Sony.

Method

The performer learned and sang the Lied 'Die Post' from Schubert's cycle *Winterreise*. Therefore, the Lied was not only previously analysed in itself, but as an integrated part of the general context of the proposed song cycle. The intended interpretation of the Lied was then communicated to the accompanying pianist. The music moments where the emotional peaks were supposed to happen were linked to the places where in the musical structure the pianist should underline the correspondent musical features. Despite the important role of the pianist in the communication of the musical structure, it was decided for the sake of the experiment that the pianist should try to stick to the singer's interpretation and to avoid any protagonist role in order not to influence the responses of the audience. She should, of course, underline the selected musical features at the previously chosen moments, but otherwise she should hold back her playing, and leave the main role of the interpretation of the song to the singer. Fifteen participants were asked to observe the performances of the Lied and to judge the musically conveyed emotional meanings. They were seated by a computer screen, and they were asked to rate in real-time feedback the emotional content, or the different emotional contents, occurring during the proposed performed interpretation by pressing emotion buttons indicated on the computer keyboard. The choice of the descriptive terms was also in this experiment limited to the four main emotions found by the musical-poetic analysis of the referred song. Each group of five screens, for technical reasons, was connected to a distribution post (Posto 1, Posto 2, and Posto 3). To each distribution post was attributed one colour: red, blue and green. So, each group of five observers' answers appeared in the graph laid out red, or blue, or green coloured. This song was performed twice, and on each occasion the audience was asked to rate the emotional meaning over the course of the performance. Each listener pressed the key according to the emotion perceived previously chosen (S-sadness; H-happiness; A-anger; F-fear). The answers were given in a real-time during the performances, and the emotion button was held down as long as the Emotion was perceived.

Experimental Procedure

After the first performance, the total of answers rated was summarised and a graph of the performance was printed (see figure 1). This graph showed a profile of all the audience selected emotions, i.e., a kind of first report of the audience's emotional meaning perception and recognition and the corresponding timing (in seconds) of the exact place within the musical structure development where the emotion has been recognised. This print out was used by the performer as a means of rectifying the extent to which the interpretation he 'thought' he had given was communicated to and detected by the audience.

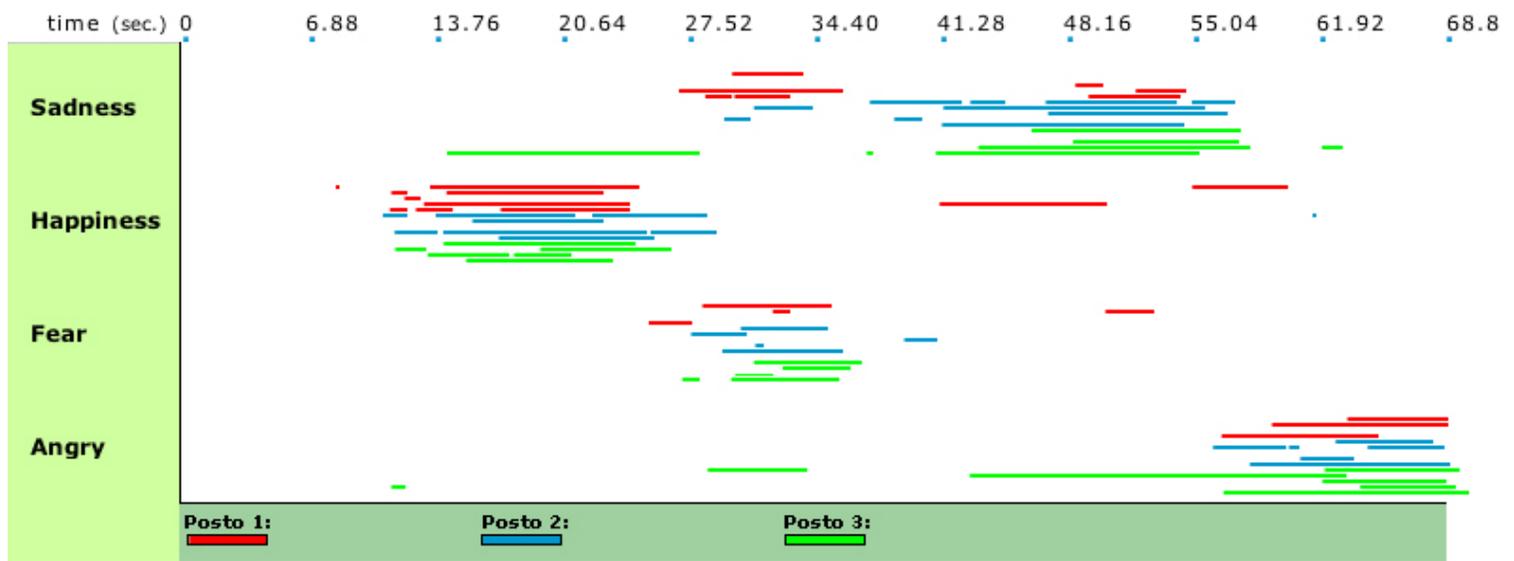


Figure 1: Profiles of the perceived emotional meaning(s) of the 1st performance of Schubert's Lied 'Die Post'

Description of the performance event

Pianist and singer started the performance with the introductory part of the piano being played on a full tale Steinway, by a professional female pianist, aged 32 (Xao Li). At bar 9 Toni started to sing in an attempt to express vocally and facially the attention required by the poetic arrival of a mailman bringing a letter with news from a beloved. The vocal and facial expressiveness developed from bar 12 to a new expressive aim, which appeared clearly and strongly with the first emotional peak at the beginning of bar 15. This emotional peak corresponded to the singer's attempt to express vocally and facially the emotion of happiness expressed in the words *mein Herz*, and with the 6th major jump (A flat 3 - F 4), being a structurally salient moment, indicating happiness as in many other cases in the music literature (see Cooke, 1959: 65). According to the performer's interpretation of Schubert's instruction at the beginning of the song *Etwas geschwind*, this moment occurred around the 20th second of the performance. According to the experimental instructions given to all the participants, the key should have been pressed as soon as an emotion was perceived and as long as the emotion remained perceptible. Here, the initial emotions of expectation and joy were indicated. Different expressive attempts followed, corresponding to the emotional peaks previously and carefully planned and rehearsed according to the performer's structural analysis of the song. These occurred around the 35th second, the 50th second, and the 70th second of the performance.

Results

The audience's recognition of the different performed emotional meanings has been again and again registered as feedback raw data on a computer laid out graph (fig. 1). The singer, after the first performance, had accessed to this graph and has carefully examined it. According to the interpretation of these first results, the singer was able to study and understand which expressive cues were missing in his performance of the song, facially as well as vocally, and those that should be improved or performed in a clear manner in order to get from his audience a better understanding of the emotional meaning he intended to communicate. The singer could also use the real time videotaped performance to observe and analyse the playback of his first performance. The missing or confusing features of his vocal and facial interpretation were identified in order to correct and improve them during the second version.

A second performance followed, then, where all the new or improved expressive cues have been added to the interpretation, and again the listeners have rated the intended emotional meaning, according to the same parameters of the first performance. The new data were printed out on a second graphic (see figure 2) and the singer was again able to compare the new results with the older ones and to check if, in the second performance, he had improved the communication of the different emotional contents of the performed song. It should be added that, though there were two different performances of the same song, the intended interpretation of the song was the same on both occasions. The only intentional change was at the level of the performance's accuracy, i.e., how the singer has been able to transmit more accurately the same emotional meanings.

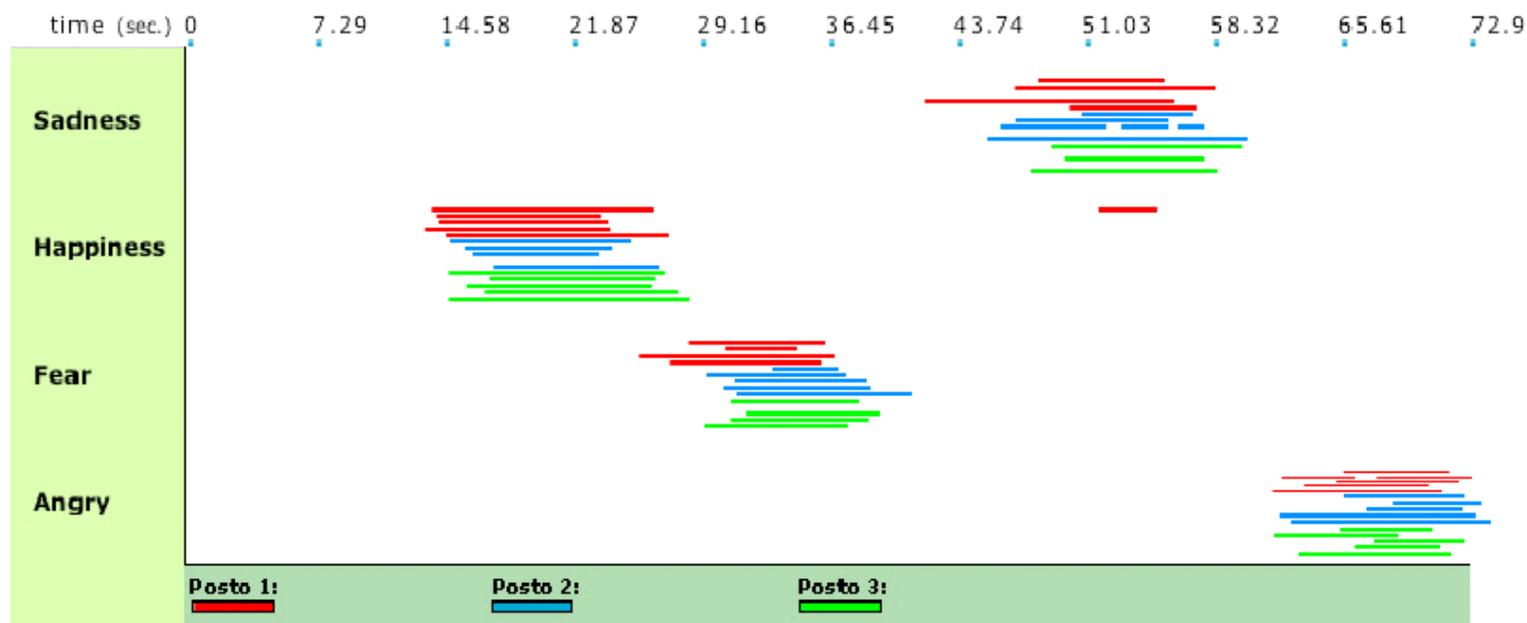


Figure 2: Profiles of the perceived emotional meaning(s) of the 2nd performance of Schubert's Lied 'Die Post'

The two previous graphs (fig. 1 and 2) show the final results of the first and second performances of Schubert's song 'Die Post'. They were printed immediately after each performance and worked as the singer's feedback from the audience perception and recognition of the different emotional meanings communicated through the two performed versions of the same intended Lied interpretation. The next two graphs (fig. 3 and 4) represent the final results summarising with more precision the experiment of the audience's emotional perception and recognition in the present study.

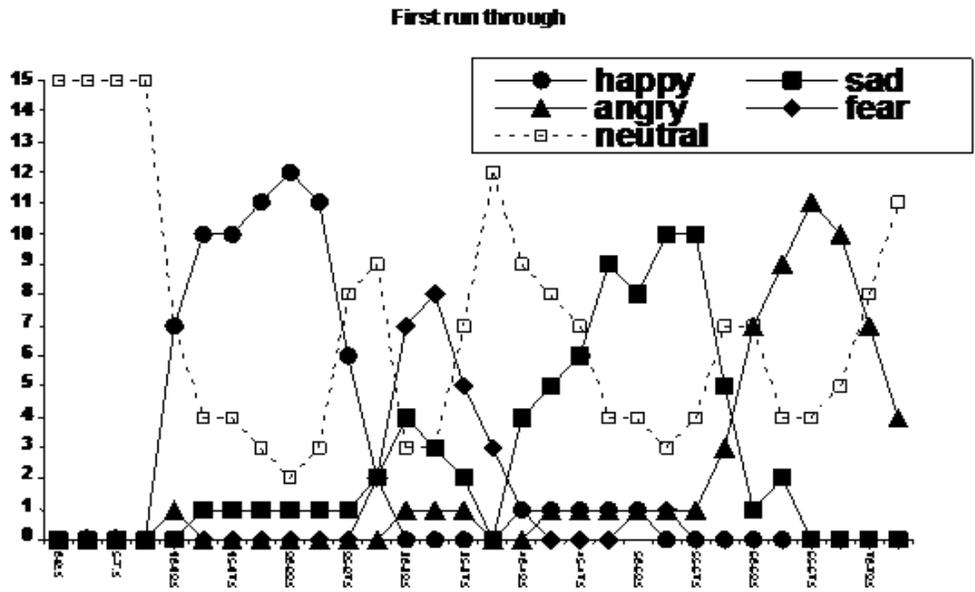


Figure 3: Graph of the perceived emotional meaning(s) of the 1st performance of the Schubert's Lied 'Die Post'

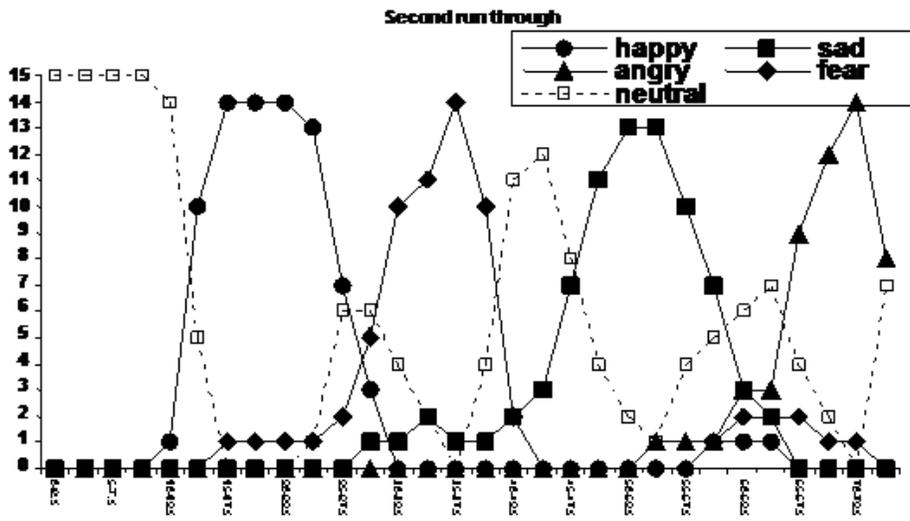


Figure 4: Graph of the perceived emotional meaning(s) of the 2nd performance of the Schubert's Lied 'Die Post'

Conclusions

With these data in hand, a series of statistical tests were undertaken to try to understand the data as fully as possible.

Comparison of overall distribution of key presses - Analysis of real time experiment

Table 1 shows that from the first to the second run through participants rated significantly fewer of the 30 moments as neutral and rated significantly more as representing fear. There were no overall differences in the mean number of presses for the other emotions (t-test).

Table 1: Mean number of moments (of 30) participants indicating each emotion

| Emotion | 1st run through | 2nd Run through |
|---------|-----------------|-----------------|
| Happy | 4.93 | 5.26 |
| Fearful | 1.86 | 4.46* |
| Sad | 5.10 | 5.10 |
| Angry | 4.06 | 3.40 |
| Neutral | 14.0 | 11.6* |

- $P < .05$ (t-test)

Accuracy of responses

Table 2 gives the mean number of the audience identified moments that correctly match the singer's intended interpretation during peak periods of each emotional expression. This is reported for each emotion separately and also an overall score is given. In addition the table shows the average number of neutral presses given during this period.

Fear is identified for significantly more time during the 4 moments in run through two than in run through one. Overall participants make significantly more correct key presses in run through two than one during these peak periods and the errors are accounted for mainly by neutral key presses - that is there were significantly more neutral presses in run through 1 than 2 during the peak emotional periods and the total correct plus neutral errors accounts for the 16 possible moments. While there was a suggestion that in run through one fear was more easily confused with anger than in run through 2 this was not confused for significantly more key moments for run through 1 than 2.

Table 2: Mean number (of 4) of correct responses during the peak emotional moments

| Emotion (key moments) | 1st run through | 2nd Run through |
|----------------------------------|-----------------|-----------------|
| Happy (7-10) | 2.93 | 3.66 |
| Fearful (13-16) | 1.53 | 3.00* |
| Sad (20-23) | 2.46 | 3.13 |
| Angry (29-32) | 2.13 | 2.86 |
| Overall correct (all 16 moments) | 9.01 | 12.67** |
| Neutral presses (all 16 moments) | 5.33 | 2.33** |

* $p < .05$

** $p < .001$ (t-test)

Discussion

By simply observing the two graphs it is apparent that the second version reveals a more defined performance of the different emotions. This indicates that the second performance has been a more successful version of the same intended interpretation of the Schubert's song. Of course, it can be suggested that every time a performance happens for the second time it becomes clearer and more meaningful for a perceiver than the first time. But, the preparation and development of the current study has revealed that this is, most of the time, not at all the

case, and that a successful performance depends really on the ability of the performer to express clearly the meaning he/she intends to communicate. In fact, in the development of the current study, it was verified that a second version of the same interpretation often caused an excess of emotional information, generating a less clear profile of the audience's responses. So, the conclusion might be that it is with the rehearsal and study of the different videotaped performances and the results of the feedback from the audience's responses for each performed version of the same song, that have shown which facial and vocal cues might be used in order for a performer be able to explore his/her personal expressive characteristics for an accurate emotional meaning communication. This study concluded, an interview with different elements of the audience suggested and corroborated the clearer profile of the second performance. The general opinion was that the facial and vocal keys were performed clearer and more accurately in the second run of the song. The singer, himself, was more convinced of the second version of his interpretation and, therefore, felt more authentic in the expressive moments of the second performance. According to the graphs summarising the final results, it can be seen that there was a clear improvement in the perception and recognition of the emotional meaning perceived in all the emotions expressed and especially the clarification of the expression of the emotion of fear.

References

- Bezooijen, R. *The characteristics and recognizability of vocal expression of emotions*. Dordrecht, The Netherlands: Foris, 1984.
- Cooke, D. *The language of music*. London: Oxford University Press, 1959.
- Davidson, J. W. Visual perception of performance manner in movements of solo musicians. *Psychology of Music* 21, no. 2 (1993): 103-13.
- Ekman, P. An argument for basic emotions. *Cognition and Emotion* 6 (1992): 169-200.
- Ekman, P. & W. V. Friesen. *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press, 1976.
- Frick, R. Communicating Emotion: the role of prosodic features. *Psychological Bulletin* 97 (1985): 412-429.
- Gabrielsson, A. & E. Lindström. 2001. The influence of musical structure on emotional expression. In *Music and Emotion: Theory and Research*, eds. P. N. Juslin & J. A. Sloboda, 223-248. Oxford University Press.
- Izard, C. E. *The emotions*. New York: Plenum Press, 1977.
- Juslin, P. N. 2001. Communicating emotion in music performance: a review and a theoretical framework. In *Music and Emotion: Theory and Research*, eds. P. N. Juslin & J. A. Sloboda, 309-337. Oxford: Oxford University Press.
- Kotlyar, G. M, & V. P. Morozov. Acoustical correlates of the emotional content of vocalised speech. *Soviet Physiology and Acoustics* 22 (1976): 208-211.
- Oatley, K. *Best laid schemes. The psychology of emotions*. Cambridge, MA: Harvard University Press, 1992.
- Plutchik, R. *The psychology and biology of emotion*. New York: Harper-Collins, 1994.
- Russel, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology* 39

(1980): 1161-78.

Salgado, A. 2001. Contribution to the understanding of some of the processes involved in the perception and recognition of emotional meaning on singers' facial and vocal expression. In *Proceedings of the 1st Meeting of the Argentina Society for the Cognitive Sciences of Music*, Buenos Aires, May 2001.

Salgado, A. *Vox Phenomena. A Psycho-philosophical Investigation of the Perception of Emotional Meaning in the Performance of Solo Singing (19th Century German Lied Repertoire)*. Unpublished Doctoral Dissertation. Sheffield University, Sheffield, 2003.

Seashore, H. G. An objective analysis of artistic singing. In *Objective analysis of musical performance*, ed. C. E. Seashore, 12-157. Iowa City, IA: University of Iowa, 1937. (*University of Iowa studies in the psychology of music* 4)

Scherer, K. R. Expression of Emotion in Voice and Singing. *Journal of Voice* 9, no. 3 (1995): 235-248.

Scherer, K. R., & H. Siegwart. Acoustic Concomitants of Emotional Expression in Operatic Singing: The Case of Lucia in *Ardi gli incensi*. *Journal of the Voice* 9, no. 3 (1995): 249-260.

Schlosberg, H. A scale for the judgement of facial expressions. *Journal of Experimental Psychology* 29 (1941): 497-510.

Schubert, E. 2001. Continuous measurement of self-report emotional response to music. In *Music and Emotion: Theory and Research*, eds. P. N. Juslin & J. A. Sloboda, 393-414. Oxford: Oxford University Press.

Tomkins, S. S. *Affect, imagery, and consciousness: The positive affects*. New York: Springer, 1962.

Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

Wundt, W. *Outlines of psychology*. Leipzig: Englemann, 1987.