frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 28 July 2011
doi: 10.3389/fpsyg.2011.00180

Authentic and play-acted vocal emotion expressions reveal acoustic differences

Rebecca Jürgens*, Kurt Hammerschmidt and Julia Fischer

Cognitive Ethology Laboratory, Leibniz Institute for Primate Research, German Primate Center, Göttingen, Germany

Edited by:
Marina A. Pavlova, Eberhard Karls University of Tübingen, Germany

Reviewed by:
Didier Grandjean, University of Geneva, Switzerland
Jochen Kaiser, Johann Wolfgang Goethe University, Germany
Angelika Lingnau, University of Trento, Italy

****Correspondence:***
Rebecca Jürgens, Cognitive Ethology Laboratory, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany.
e-mail: rjuergens@dpz.eu

Play-acted emotional expressions are a frequent aspect in our life, ranging from deception to theater, film, and radio drama, to emotion research. To date, however, it remained unclear whether play-acted emotions correspond to spontaneous emotion expressions. To test whether acting influences the vocal expression of emotion, we compared radio sequences of naturally occurring emotions to actors' portrayals. It was hypothesized that play-acted expressions were performed in a more stereotyped and aroused fashion. Our results demonstrate that speech segments extracted from play-acted and authentic expressions differ in their voice quality. Additionally, the play-acted speech tokens revealed a more variable F_0-contour. Despite these differences, the results did not support the hypothesis that the variation was due to changes in arousal. This analysis revealed that differences in perception of play-acted and authentic emotional stimuli reported previously cannot simply be attributed to differences in arousal, but by slight and implicitly perceptible differences in encoding.

Keywords: emotions, vocal expressions, authenticity, acting, acoustic analysis

INTRODUCTION

Emotional expressions are an important aspect in our daily communication. Different emotions are characterized by special expression patterns, such as facial expressions and vocal characteristics, but also body postures and movement patterns. Humans can recognize these patterns quite reliably, even across cultures (reviewed by Cowie et al., 2001), with some differences with regard to the emotion and the modality in which it is conveyed. To date, research in this field concentrated mainly on actors' portrayals of emotions (for facial expression: e.g., Ekman et al., 1969; for vocal expression: e.g., Banse and Scherer, 1996). Play-acted expressions were assumed to be at least closely related to authentic ones; otherwise listeners would not be able to recognize the emotional content of these portrayals (Juslin and Laukka, 2001; Scherer, 2003). The transitions between authentic and play-acted emotional expressions appear seamless and whether a differentiation between these two encoding conditions is reasonable at all is under discussion (Scherer, 2003; Klein, 2010). This view is supported by the fact that portrayals may be based on felt emotions, as generated by the "method acting" technique (Strasberg, 1987) and on the observation that the expressions of emotions are influenced by display rules (Ekman and Oster, 1979) and are thus also an outcome of acting. However, emotions are accompanied by physiological reactions (Kreibig, 2010) that are not under full voluntary control and that influence expression (Scherer, 1986) in a way that is thought to be imitated only with difficulty (Juslin and Laukka, 2001). Until now, the exact relation between the intentional communication displays represented by portrayals (cf. Elfenbein et al., 2010) and authentic, spontaneous emotional expressions that are accompanied by corresponding underlying emotions is, for the most part, still ambiguous. It is the aim of this study to shed more light on this discussion in regard to vocal expressions of emotions.

For facial expressions, authentic emotions are distinguishable from play-acted ones by symmetric and simultaneously occurring muscle movements (Ekman and O'Sullivan, 2006). Less attention has been given to vocal expressions, although the voice is of particular interest in this context as it is strongly linked to changes in physiology (Scherer, 1986; Owren and Bachorowski, 2007). So far, research has shown that play-acted emotional expressions in speech seem to be at least more stereotypical and overemphasized compared to authentic ones (e.g., Barkhuysen et al., 2007; Laukka et al., 2007), which would be in line with the signal character of portrayals (Scherer, 1986, 2003). Barkhuysen et al. (2007) demonstrated for example that listeners who were confronted with recordings of people expressing induced or acted emotions judged the authentic emotions as less extreme than the play-acted ones. Similar results were obtained from Laukka et al. (2007), who compared the perception of induced and re-enacted emotional expressions in professional actors. Recently, a study by Scheiner and Fischer (2011), which analyzed cross-culturally the effect of authenticity on the discrimination of different emotions on the basis of naturally occurring emotions, revealed that raters were poor at explicitly distinguishing between play-acted and authentic emotions (see Audibert et al., 2008). Surprisingly though, the source of the recording had a significant effect on emotion recognition, suggesting that the listeners did pick up some differences in the stimuli.

The question of which differences lie in the stimuli themselves and were responsible for the different perception of play-acted and authentic expressions is still not fully answered, especially with regard to naturally occurring emotions. To our knowledge, the first direct acoustic comparison of authentic and play-acted vocal expressions was conducted by Williams and Stevens (1972). They compared the voice of the radio announcer reporting the crash of the Hindenburg airship with an actor re-enacting the scene. Their results showed that

the arousal-related influences on the fundamental frequency (F_0) and the F_0-variability were more pronounced for the actor's voice than for the original speaker. More recently, Audibert et al. (2010) conducted a pilot study of a direct acoustic comparison using the emotional stimuli from the study by Laukka et al. (2007). The induced and the re-enacted emotional expressions of anxiety, irritation, and satisfaction of four actors were compared with the result that the play-acted emotions are characterized by a higher F_0, a lower second formant (F2) and a higher F_0-variability. The higher, more variable F_0 correlates highly with activation and intensity (Laukka et al., 2005). Finally, recent acoustic analyses on authentic expressions, that did not include direct comparisons, also revealed weaker acoustic differences between the emotions for the authentic expressions compared to studies on portrayals (Laukka et al., 2011).

Despite these findings, a detailed acoustic comparison of authentic and play-acted expressions with a reliable sample size was still missing. To reveal whether and by which acoustic structures play-acted expressions can be differentiated from authentic ones, we conducted an acoustic analysis using emotional radio recordings and play-acted equivalents by professional actors (Scheiner and Fischer, 2011). This study followed an explorative approach using a multivariate acoustic analysis to obtain a detailed description of the stimuli. Based on the previous studies (Williams and Stevens, 1972; Barkhuysen et al., 2007), we hypothesized that play-acted emotions are more aroused and more intense than authentic ones. We therefore predicted at least a higher F_0, more vocal perturbation, higher formants, smaller bandwidths of formants, more energy in the higher frequency regions, a faster speed of speech, and a higher F_0-variability for play-acted expressions, as these acoustic parameters correlate strongly with arousal (see Laukka et al., 2005; Owren and Bachorowski, 2007). On the suprasegmental level, we compared the speed of speech and the F_0-variability, but not the intensity, as the uncontrolled recording conditions of the radio sequences did not allow such comparisons. Voice quality parameters as well as the fundamental frequency were analyzed using cut-out vowels as they have comparable and stable acoustic characteristics (Bachorowski and Owren, 1995).

MATERIALS AND METHODS

RECORDINGS

The recordings of the authentic emotional expressions were collected from a radio archive in Hamburg, Germany. Interview sequences were selected in which people experienced one of the four emotions "anger," "fear," "sadness," or "joy" or reported emotionally about situations in the past, as for example, parents speaking about the death of their children, people getting angry about injustice or being afraid of a current threat. These sequences were transcribed and the context of the recorded situation was noted. The language of all sequences was German. Short speech fragments that did not contain emotional keywords were cut-out from the complete sequences. Finally, 80 speech fragments by 78 speakers were selected and saved as wave files (see Appendix I for example transcripts). Two speakers contributed two stimuli to the set, one man producing two "sadness" stimuli and one woman producing one "fear" and one "joy" stimulus. One "fear" stimulus was rejected from the acoustic analysis because the recording was compressed such that the frequency range was too low for comparable measurements. The remaining 79 stimuli spoken by 77 speakers had an average duration of 1.859 s (range 0.343–5.491 s). An overview over the number of speakers for all conditions is given in **Table 1**.

The entire sequences were re-enacted by actors who were recruited in Berlin, Hanover, and Göttingen (21 male and 21 females; 31 professional actors, 10 drama students, 1 professional singer). The actors were provided with the transcript, the context of the situation, and the respective emotion. They could perform the sequences multiple times and selected the version they preferred most. The recordings were made with a Marantz Professional Portable Solid State Recorder (Marantz, Kanagawa, Japan, 44.1 kHz sampling rate, 16 bit sampling depth) and a Sennheiser directional microphone (Sennheiser, Wedemark, Germany, K6 power module and ME64 recording head). Almost half of the authentic speech tokens (35 out of 80) were made outdoors and varied in terms of their surrounding noise. To avoid a discrimination effect due to the background noise, a comparable number (30) of randomly

Table 1 | Number of vowels and speakers per condition.

		Vowels[a]						Speaker vowels[b]						Speaker fragments[c]	
		Authentic			Play-acted			Authentic			Play-acted			Authentic	Play-acted
		a,	e,	i	a,	e,	i	a,	e,	i	a,	e,	i		
Male	Fear	16,	16,	8	16,	14,	15	7,	7,	5	4,	4,	3	8	5
	Anger	25,	20,	21	14,	16,	10	12,	12,	10	6,	6,	5	12	6
	Joy	19,	24,	16	17,	25,	12	10,	10,	10	5,	6,	6	10	6
	Sadness	24,	18,	13	15,	18,	12	9,	9,	7	4,	5,	5	10	5
Female	Fear	20,	19,	9	13,	17,	7	9,	9,	5	5,	5,	4	9	5
	Anger	21,	20,	16	14,	18,	6	10,	10,	9	6,	7,	4	10	6
	Joy	23,	27,	14	22,	18,	6	8,	8,	8	4,	5,	4	10	5
	Sadness	18,	17,	13	11,	14,	3	9,	9,	9	4,	5,	3	10	5

[a] *Given are the vowels with a tonality in more than 10% of all time segments and with correctly calculated first and second formants that were used for the averaging over the speakers. The total number of vowels is 770.*

[b] *The different vowels of one condition were spoken by the same speakers, while the speakers are independent across emotion, gender, and source condition.*

[c] *Speakers are not totally independent of each other, as some contribute more than one stimulus to the set.*

selected re-enactments was also recorded outside. The actors spoke mostly two sequences of the same emotion, with the exception of one male actor that contributed three "fear" stimuli, one man that contributed three "sadness" stimuli and one actress that contributed three "joy" stimuli to the set. One actor contributed one "fear" and one "sadness" stimulus to the set. The speech fragments were cut and saved as wave files.

ACOUSTIC ANALYSIS

Acoustic structure of vowels

Due to a lack of "o" (/ɔ/) and "u" (/ʊ/) in the stimuli only "a" (/a/), "e" (/ɛ/) and "i" (/I/) were cut-out of the fragments with 0.5 s of silence at the beginning and the end using the Avisoft-SASLab-Pro Version 4.52 (Avisoft Bioacoustics, Berlin, Germany). Nine hundred ninety-seven vowels from the different speakers were selected that had a mean duration of 0.082 s (SD = 0.041 s, range = 0.022–0.393 s).

The formant analysis was done using Praat 5.1.11 (Boersma and Weenink, 2009) in combination with the quantify formants script of the GSU Praat tools 1.9 (Owren, 2008), a script package that allows batch processing during measurements. Before measuring the formant locations and the bandwidths, the stimuli were pre-emphasized with 6 db per octave beginning with 50 Hz, to amplify the higher frequencies. The formants were calculated using an algorithm by Burg (Press et al., 1992) with the following settings: maximal number of formants: 5, maximal value of formants: 5000 Hz for male speakers, 5500 Hz for female speakers, window length: 0.025 s, window placement: around peak amplitude. A pretest with a selection of files was performed to identify the appropriate settings. During the calculations, all measurements were checked visually using broadband spectrograms with overlaid formant structures generated in the Praat sound editor. For 86.66% of all measurements (864 out of 997 vowels) the first two formants were calculated correctly. The high number of miscalculations, mostly concerning the vowel "i," is explained by the bad quality of the vowels, which were partly quite short, noisy, or poorly articulated.

The parameters related to the F_0, to the energy distribution and the vocal perturbation were measured using LMA ("Lautmusteranalyse"), a program that analyses spectrograms (developed by Hammerschmidt) and that calculates two different sets of parameters. The first calculation included only tonal segments in the calculation and measured parameters related to the F_0 and the vocal perturbation (dubbed tonal calculation). The second calculation (dubbed general calculation) measured parameters in tonal as well as in noisy segments and included the parameters for energy distribution (see Schrader and Hammerschmidt, 1997 for description of the algorithms; Hammerschmidt and Jürgens, 2007). Spectrograms were created using Avisoft-SASLab to conduct the LMA analyses. For the tonal calculation of the vowels "a" and "e," a FFT (1024 points) with a sampling frequency of 5500 Hz, Hamming window and 98.43% overlap was performed that generated spectrograms with a frequency range of 2750 Hz, a frequency resolution of 5 Hz and a time resolution of 3 ms. As the vowel "i" is characterized by fewer intense harmonics in the lower regions, a wider frequency range was required to permit the detection of tonality. For the vowel "i," we conducted a FFT with a sampling frequency of 7200 Hz, generating spectrograms with a frequency range of 3600 Hz, a frequency resolution of 7 Hz and a time resolution of 2 ms. The spectrograms were then analyzed using LMA with the help of an interactive harmonic cursor to conduct the tonal calculation. In each spectrum the F_0 region was marked to predefine the area where the algorithm was to search for the F_0. This was helpful as, due to the background noise, the F_0 was in part not clearly defined at the start and end points, which would have led to miscalculation without predefining. Before the final calculation was executed, a test was performed to control the matching of the visual spectrum and the calculation. The tonal parameters are only reliable in cases in which tonality can be detected in more than 10% of the time segments. This was the case for 89.5% of all vowels (903 out of 997). For 781 out of the 997 (78.3%) vowels, both the formants and the tonal parameters could be analyzed adequately. To perform the general calculation a FFT (1024 point) with a sampling frequency of 8000 Hz, Hamming window and 98.43% overlap was conducted to enlarge the frequency range for measurements of energy distribution. This FFT resulted in spectrograms with a frequency range of 4000 Hz, a frequency resolution of 8 Hz and a time resolution of 2 ms. The spectrograms were then analyzed with the general calculation to calculate the second set of parameters. To reduce influence of noise on the measurements, all LMA analyses were conducted using a cut-off frequency of 50 Hz and start and end thresholds of 10 which led to a rejection of all time segments with amplitudes lower than 10% of the maximal amplitude of the utterance.

Suprasegmental level

Two different parameters concerning the speed of speech were measured using the speech fragments. While the speech rate is defined as the duration of utterances including pause intervals (Jacewicz et al., 2009), the articulation rate excluded pauses (Quené, 2008). The total duration of the speech fragments were measured with the Avisoft-SASLab by measuring the distance between the first and the last visible articulation in the envelope. The speech rate was then obtained by dividing the syllables of the speech fragments through the total duration. For the articulation rate, all sections without audible articulation were measured manually using an FFT with the following settings: FFT length: 1024 points, sampling frequency: 5.5 kHz, Hamming window and 98.43% overlap, resulting in spectrograms with a frequency resolution of 5 Hz and a time resolution of 2.9 ms. The articulation rate was then calculated by dividing the syllables through the duration of audible articulation.

The variability of the F_0 on the basis of the speech fragments was analyzed by measuring the F_0 in intervals of 0.2 s by hand using the Avisoft-SASLab-Pro Free reticule cursor. For this purpose, spectrograms were generated (sampling frequency of 2.2 kHz, Hamming window, and 98.43% overlap) with a 1.1-kHz frequency range, a time resolution of 7 ms, and a frequency resolution of 2 Hz. The SD of the F_0 measurements (F_0 SD) was than calculated and used as the parameter for F_0-variability.

STATISTICAL ANALYSIS

Acoustic structure of vowels

To extract a small set of uncorrelated factors out of the large set of parameters calculated from the vowels, a principal component analysis with varimax rotation and Kaiser normalization (KMO = 0.864) was conducted for all vowels for which LMA

detected a tonality of more than 10% and for which at least the first two formants were calculated correctly ($N = 781$). The analysis resulted in 13 factors with an Eigenvalue greater than 1 that explained 76.7% of the variance. The interpretation of each factor and its explained variance are summarized in **Table 2** (see Appendix II for the description of all parameters with high factor loadings).

We tested the normal distribution of the factors using a Kolmogorov–Smirnov-Test, which indicated a normal distribution for all factors ($z \leq 1.157, p \geq 0.137$), with the exception of factor 9 for the cells vowel_i-female-sadness-authentic ($z = 1.421$, $p = 0.035$), and vowel_e-male-anger-authentic ($z = 1.383$, $p = 0.044$). In light of the large number of comparisons, these effects can be considered negligible, and they would be rendered non-significant after correction for multiple testing. The Levene-test (based on median) for homogeneity of variance revealed that variance was mainly homogeneous in case of EMOTION ($2.596 \geq W \geq 0.041$, $0.056 \leq p \leq 0.989$), SOURCE, and GENDER ($3.91 \geq W \geq 0.005$, $0.051 \leq p \leq 0.944$) with the following exceptions: Factor 12 in the EMOTION condition for vowel e ($W = 4.52$, $p = 0.005$), Factor 10 in the SOURCE condition for vowel a and e, and Factor 11 in the GENDER condition for vowel a and i ($W \geq 4.719$, $p \leq 0.032$).

The global hypothesis of whether the acoustic structure of the stimuli was influenced by the conditions was tested by using a multivariate General Linear Model (multivariate GLM, PASW 17). The vowels "a," "e," and "i" differ in their formant structure and their energy distribution and were therefore calculated separately. As the speakers contributed more than one of each vowel to the set all factor values were averaged over the speakers so that for each speaker, vowel and emotion only one value per factor was used. Two actors contributed vowels to two different emotions (one female: "fear" and "joy," one male: "fear" and "sadness"). As there were fewer "fear" stimuli, the "joy," and "sadness" stimuli were left out of the further analysis to make all samples independent of each other, which reduced the final vowel set to 770 (see **Table 1** for an overview over the number of speakers and vowels for all conditions). The averaged factor values were than analyzed using the multivariate GLM in terms of GENDER, EMOTION, and SOURCE. In cases in which the multivariate analysis resulted in differences across conditions, the factors were tested separately using Linear Mixed Models (LMMs, PASW 17) to look for differences in the factors across the conditions that showed some influence on the global acoustic structure. For the univariate LMMs, we conducted transformations to obtain homogeneity of variance in the respective factors. Values of Factor 11 in the GENDER condition (vowel a and i) and values of Factor 10 in the SOURCE condition (vowel a) was transformed using the cube transformation, while the values of vowel e for Factor 10 in the SOURCE condition was transformed using the log transformation ($W \leq 3.183$, $p \geq 0.078$). Again, the analysis was separated by vowels.

Suprasegmental level

The speech rate and the articulation rate of the speech fragments were tested for influence of GENDER, EMOTION, and SOURCE also by using LMMs (PASW 17), additionally SPEAKER was added as a random factor. A Bonferroni correction was used for the *post hoc* tests. The speech segments were taken from different parts of sentences and had different lengths and stress patterns, all of which can influence intonation (Botinis et al., 2001). Hence, a

Table 2 | Influence of the different conditions on the factors for the vowels "a," "e," and "i."

	Factors	Explained variance (%)	SOURCE		GENDER		
			a	e	a	e	i
F1	Peak frequency (PF), first quartile of distribution of frequency amplitudes (dfa 1)	19.90	0.047 ↓	–	–	–	0.027 ↑
F2	Frequency range, second quartile of DFA (dfa 2), third quartile of DFA (dfa3)	14.20	–	–	–	–	–
F3	Trend and modulation of the PF	7.80	–	–	–	–	–
F4	Fundamental frequency	6.70	–	–	0.000 ↑	0.000 ↑	0.000 ↑
F5	Percentage of tonal segments	4.50	–	0.010 ↓	–	–	–
F6	Bandwidth of the first formant (BWF1)	3.80	0.000 ↑	–	–	–	–
F7	Amplitude ratio between first harmonic and F_0 (amprat2), and between third and second harmonic (amprat3)	3.40	0.001 ↓	0.004 ↓	–	0.002 ↓	0.002 ↓
F8	Harmonic-to-noise-ratio (HNR)	3.40	–	–	–	0.006 ↑	0.011 ↑
F9	Jitter	3.00	–	–	–	–	–
F10	Location of maximum frequency amplitudes	2.90	–	–	–	–	–
F11	Shimmer	2.50	–	0.042 ↑	–	–	–
F12	Location of the minimum PF, location of maximum correlation coefficient of successive time segments	2.40	–	–	–	–	–
F13	Correlation coefficient of successive time segments	2.20	–	–	–	–	–

Given are the interpretations of the factors, their explained variance, and the p-values of the LMMs. Upward directed arrows indicate an increased value from authentic to play-acted (SOURCE) or from male to female (GENDER), downward directed arrows indicate a decreased value.

comparison of the F_0-variability was only possible for the respective pairs (authentic speech stimulus and play-acted equivalent) in terms of authenticity. The influence of emotion and their interaction with authenticity could not be tested. For the analysis, a repeated LMM (PASW 18) with diagonal covariance structure was used that tested the F_0 SD for all stimulus pairs with SOURCE as repeated factor and EMOTION and SPEAKER as a random factor.

RESULTS

ACOUSTIC STRUCTURE OF VOWELS

The multivariate analysis of the 13 factors revealed global differences in the GENDER (Pillai's – Trace = 0.496, $F = 6.348$, $p = 0.000$) and SOURCE factor (Pillai's – Trace = 0.280, $F = 2.510$, $p = 0.006$) for vowel "a." Surprisingly, no global differences were found between the emotions (Pillai's – Trace = 0.473, $F = 1.238$, $p = 0.169$). The results for the vowel "e" were similar (GENDER: Pillai's – Trace = 0.532, $F = 7.787$, $p = 0.000$; SOURCE: Pillai's – Trace = 0.228, $F = 2.021$, $p = 0.028$; EMOTION: Pillai's – Trace = 0.393, $F = 1.055$, $p = 0.388$). The fact that no interactions between any of the conditions could be identified ("a": $0.190 \leq$ Pillai's – Trace ≤ 0.424, $0.973 \leq F \leq 1.513$, $p \geq 0.130$; "e": $0.085 \leq$ Pillai's – Trace ≤ 0.453, $0.633 \leq F \leq 1.244$, $p \geq 0.162$), indicated that the differences between play-acted and authentic emotions were independent of the emotional expression. For the vowel "i," only the gender influenced the acoustic parameters (GENDER: Pillai's – Trace = 0.520, $F = 5.750$, $p = 0.000$; SOURCE: Pillai's – Trace = 0.159, $F = 1.002$, $p = 0.459$; EMOTION: Pillai's – Trace = 0.535, $F = 1.185$, $p = 0.224$; interactions: $0.225 \leq$ Pillai's – Trace ≤ 0.548, $0.996 \leq F \leq 1.544$, $p \geq 0.124$). The lack of significant differences for vowel "i" can be explained by a lower statistical power as a result of a smaller sample size, as most miscalculations during measurements occurred for the vowel "i."

The subsequent LMMs demonstrated that authentic and play-acted stimuli differed in 5 of the 13 factors (**Table 2**; **Figure 1**). Though the LMMs did not result in the same significant differences for both vowels, the figure shows that at least the tendencies in which the parameters differ in the SOURCE condition were similar. The most consistent differences were found for Factor 7 as these were detected in both vowels. Factor 7 is most strongly associated with the amplitude ratios between the third and the first harmonic, and between the second harmonic and the F_0. While a value of 1 reflects an equal intensity of both frequency bands, lower values, as found for the play-acted stimuli, indicate more dominant lower frequencies (**Figure 2**). In regard of the factor loadings, play-acted emotional utterances were characterized by lower peak frequencies and more energy in the lower frequency regions (F1), less tonality (F5), broader bandwidths of the first formants (F6), more dominant lower harmonics (F7) and higher shimmer values (F11) compared to authentic ones.

The acoustic structure of the vowels was furthermore influenced by GENDER, which was not surprising. The LMMs (**Table 2**) demonstrated a higher peak frequency (F1), higher F_0 (F4), more dominant lower harmonics (F7) and a higher harmonic-to-noise-ratio (HNR) indicating clearer speech (F8) for female speakers.

SUPRASEGMENTAL LEVEL

The LMMs showed that neither the speech rate nor the articulation rate were influenced by GENDER (speech rate: $F = 0.405$, $p = 0.526$, articulation rate: $F = 1.814$, $p = 0.18$) or SOURCE (speech rate: $F = 0.024$, $p = 0.875$, articulation rate: $F = 0.078$, $p = 0.78$). Furthermore, the articulation rate did not vary between the emotions ($F = 1.228$, $p = 0.302$). The only difference that was found concerning the speed of speech, was an effect of the emotions on the speech rate ($F = 3.703$, $p = 0.013$). The *post hoc*-test with Bonferroni correction showed that "anger" (mean = 6 syllables/s, SD = 1.6 syll/s) was spoken faster than "sadness" (mean = 5 syll/s, SD = 1.9 syll/s; $p = 0.01$). As the speech rate differed from the articulation rate in terms of the included pauses, "anger" stimuli were characterized by fewer pauses than "sadness" stimuli, although they were articulated with the same rate. No interactions between the conditions were found ($0.156 \leq Z \leq 1.261$, $p \geq 0.284$). SPEAKER could not explain any variance. The paired LMM demonstrated that the play-acted stimuli were generally spoken with a higher F_0 SD than their authentic counterparts (estimated difference = 7.8, $F = 6.325$, $p = 0.013$) which revealed a higher variability of the F_0-contour for the play-acted speech tokens (**Figure 3**). EMOTION (Wald $Z = 0.986$, $p = 0.324$) and SPEAKER (Wald $Z = 1.195$, $p = 0.232$) did not contribute much to the model.

DISCUSSION

AUTHENTICITY RELATED DIFFERENCES

This study revealed an influence of acting on the F_0-variability and on the acoustic structure of vowels. Play-acted expressions were characterized by a higher amplitude of the lower harmonics, by broader bandwidths of the first formant, lower peak frequencies, more amplitude fluctuations (higher shimmer values), less tonality, and by a higher overall variability of the F_0-contour compared to authentic expressions. With the exception of the F_0-variability, other parameters that are strongly associated with arousal, like the mean F_0, the HNR, or the speech rate (Laukka et al., 2005; Owren and Bachorowski, 2007), were not affected by the encoding condition contrary to our initial hypothesis. Furthermore, while aroused speech is connected to narrower bandwidths of formants due to a decreased level of salivation (Scherer, 1986; Laukka et al., 2005) and high peak frequencies (Hammerschmidt and Jürgens, 2007), we observed the opposite. These results demonstrate that the differences between authentic and play-acted emotional expressions cannot solely be explained by arousal. As we could neither detect an overemphasized encoding for play-acted expressions nor any other interactions between emotion and source conditions, the effect of acting seems to be independent of the emotional expression and support the view that the encoding of play-acted and authentic emotional stimuli differs in some way.

As other arousal-related parameters were not affected, the differences in the F_0-variability, also found by Williams and Stevens (1972) and Audibert et al. (2010), might be caused by more strongly stressed and more variable speech during acting. The differences found in the vowel structure might be related to a higher degree of glottal leakage in the actors' voices that resulted in a more breathy speech. Hanson and Chuang (1999) summarized that breathy voices were characterized by more intense fundamental frequencies, broader bandwidths of the first formants and aspiration noise in the region around the third formant. Differences in the aspiration noise could not be detected in our stimulus set (unpublished data), but as the measurements of HNR in the higher frequency

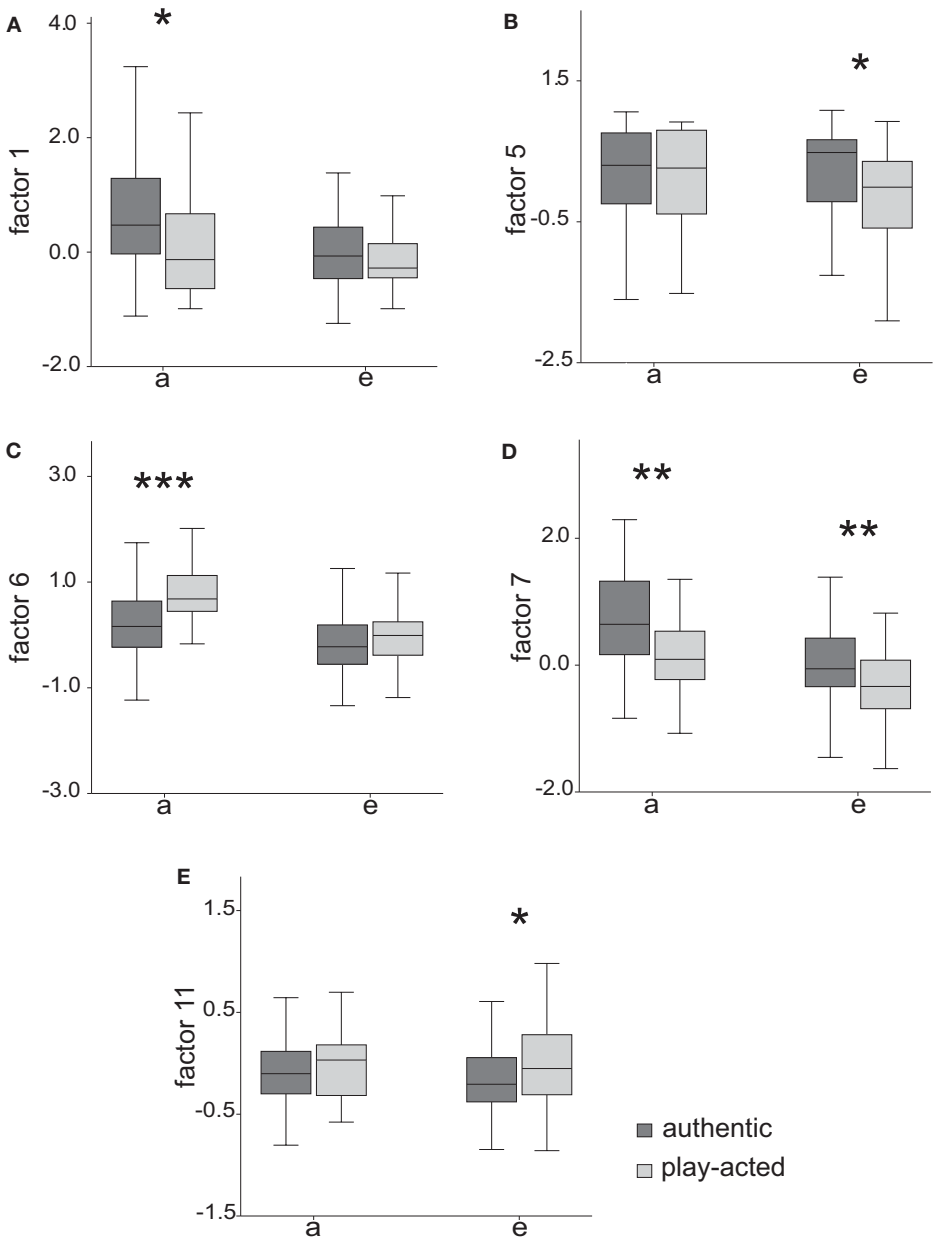

FIGURE 1 | Differences in the factor loadings between authentic and play-acted speech tokens. Given are the median, the lower, and the upper quartile. Whiskers represent the values within the 1.5 interquartile range. The figures **(A)** to **(E)** represent the 5 factors that varied in the SOURCE condition. The asterisks mark the significant levels of the differences found by the LMMs (*<0.05, **<0.01, ***<0.001).

regions might be less reliable due to the weak signal intensity of the frequency band, the presence of aspiration noise cannot be ruled out completely.

The differences found for authentic and play-acted expressions might be explained either by the acting process *per se* or by the characteristics of the actors' speech. They are possibly not specifically related to the emotion expression but might be present in neutral speech as well. A monotonous intonation is perceived as tiresome and uninteresting (Botinis et al., 2001) and it is possible that actors are taught to speak more dynamically to entertain their audience. It is also known that voice training has an effect on voice quality (Master et al., 2008) as, for example, actors have a special energy peak around 3.5 kHz (Nawka et al., 1997) called the actors formant. Furthermore, Master et al. (2008) found acoustic cues that indicated more favorable glottal adjustment and a faster glottal closing in actors. While our results, do not point in a similar direction, a comparison with subject without acting skills and neutral speech would nevertheless be helpful to identify the effect of actors' speech.

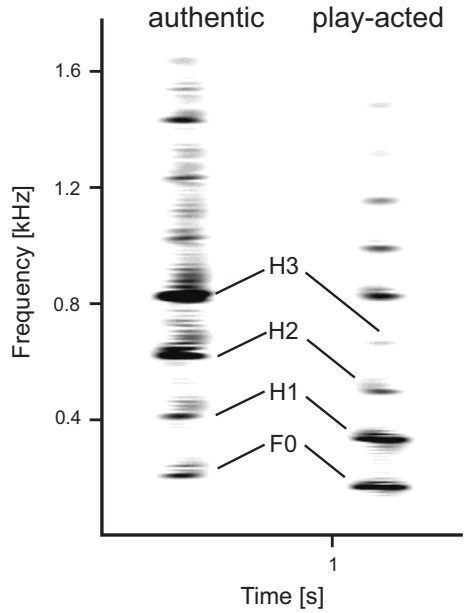

FIGURE 2 | Differences in the energy distribution of the lower frequency bands between authentic and play-acted utterances. The authentic stimulus possesses high amplitude ratios between the second harmonic and the F_0 (amprat 2 = 3) and between third and first harmonic (amprat 3 = 2.71), while the play-acted stimulus is characterized by lower values (amprat 2 = 0.28, amprat 3 = 0.3). The differences in the F_0 positions are due to individual differences. Given are the FFT spectrograms of one authentic and one play-acted female spoken "a" with a sampling rate of 2.7 kHz (Avisoft-SASLabPro).

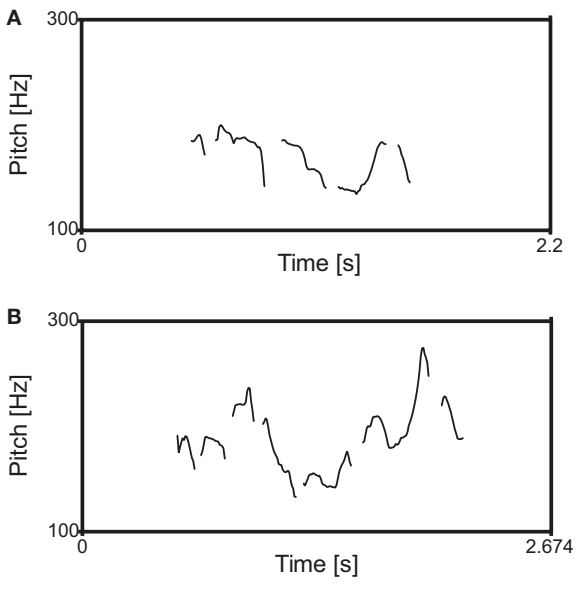

FIGURE 3 | F_0-contour of one authentic (A) and one play-acted (B) anger stimulus (text fragment: "[Polizei]präsenz is und so, was hier los ist," translated: "[police] presence is and such, what is going on here") that demonstrates the higher F_0-variability of the play-acted stimuli. The pitch-contour was generated using Praat.

In contrast to other studies, we did not find overemphasized and stereotypical acted expressions. The recognition study by Barkhuysen et al. (2007) was based on the Velten mood induction technique (Velten, 1968), while the stimulus material used by Laukka et al. (2007) was generated using a language training program. Reading sentences according to the Velten mood induction technique might generate emotions in the participant (Westermann et al., 1996), but it is questionable whether they are intense enough to produce strong emotional expressions. In regard to computer programs, one cannot exclude the possibilities that the subjects were emotionally not involved or that the generated emotions were partly masked or repressed, even unconsciously, due to display rules (Ekman and Oster, 1979), as they were generated in laboratory surroundings with participants knowing they were under observation. The low intensity of emotion induction via computer games was demonstrated by Kappas and Polikova (2008). Although the procedure of using induced emotions has its advantages, it is not surprising that less intense emotional expressions were detected in comparison to acted ones. On the other hand, the procedure of asking actors to express one emotion in one special utterance is well suited to produce overemphasized, stereotypical expressions. The differences in the intensity levels seem therefore to be related to the study design and not to acting *per se*. By providing the actors with long speech sequences and with contexts, stimuli were created that might be nearer to the more naturalistic acted emotions (see also Goudbeek and Scherer, 2010). As no differences in the intensity were present across the encoding conditions in our study, we were able to detect the effects that lie solely in the acting itself.

The lack of evidence for a more stereotypical encoding of emotions by actors fits with the observation that listeners did not recognize the play-acted expressions more accurately than the authentic ones, revealed in the recognition study by Scheiner and Fischer (2011) that uses the same speech material. In this recognition study, the subjects were not able to discriminate authentic from play-acted stimuli. However, the recognition experiment revealed an influence of the encoding condition on the emotion judgment: listeners rated anger more accurately when play-acted and sadness more accurately when authentic. Thus, the subtle acoustic differences uncovered in the present study implicitly affected the emotion recognition of the raters, at least in two of the emotions tested. As the analysis of the vowels did not reveal acoustic differences in relation to emotion, the interaction between emotion recognition and recording condition could be caused by the more variable F_0-contour of play-acted stimuli, since a variable F_0-contour is related to aroused expressions like anger (Juslin and Laukka, 2003). Whether the acoustic differences affect the emotions judgment directly is be tested in further studies.

The results of this study should be seen under the limitation that stimuli were used that were partly based on emotional memories. It can be asked how emotional they really are. Furthermore, as Scherer (2003) mentioned, emotional stimuli taken from the media might be affected by social acting. Even though the stimulus set was composed of situations in which acted self-portrayal was thought to be low, as opposed to stimuli taken from talk shows, the effect of social acting can never be completely excluded. Another limitation

results from the fact that the authentic and the play-acted speech tokens were based on a different amount of speakers, what might influence the results due sample composition.

GENDER AND EMOTION-RELATED DIFFERENCES

In addition to the effect of authenticity, we detected an influence of gender on the acoustic structure of vowels. This was not surprising as female voices can be differentiated easily from male voices (Lass et al., 1976) and a number of studies have already characterized the acoustic differences (e.g., Titze, 1989). Our results, higher HNR and higher F_0 that are more intense than the overlying harmonics, correspond to previous results for female speakers (see Hammerschmidt and Jürgens, 2007), indicating that the analysis of the cut vowels produced valid and comparable results.

While there was an effect on speech rate, we did not identify any influence of emotion on the acoustic structure of the vowels. In line with previous studies, sad expressions were spoken with more pauses than angry expressions (e.g., Sobin and Alpert, 1999). Differences between other emotions were not found. Due to the non-standardized sentences, an influence of emotion on the F_0-variability could not be conducted. As a large number of studies identified acoustic cues that differentiate the sound structure of emotional utterances (review: Juslin and Laukka, 2003), it was surprising that we could not. There are three possible, not mutually exclusive explanations for the lack of emotion-related acoustic cues. First, the multivariate statistical analysis is quite conservative and rejects differences when they lie solely in a small number of parameters. In combination with the factor analysis this might lead to a serious loss of information. Second, the analyzed speech segments (vowels) were quite short. Even though Bachorowski and Owren (1995) were able to detect an influence of positive and negative emotions on single acoustic cues in comparable speech segments, other studies that found emotional differences in vowels cut from running speech dealt with vowels nevertheless twice as long as 0.08 s (Leinonen et al., 1997; Waaramaa et al., 2010). Further studies should analyze longer segments and should concentrate specifically on the prosody parameters, as the differences in the F_0-variability for authenticity and in the speech rate for emotion differentiation demonstrated the importance of the suprasegmental level. The fact that gender and authenticity could nevertheless be differentiated in our study emphasized their strong effect on the acoustic structure. Third, the lack of emotion-related cues might be due to the quality of the stimuli. The recognition experiment mentioned before (Scheiner and Fischer, 2011) demonstrated that listeners could recognize the emotions only in 40% of all cases (mean across listeners of Germany, Romania, and Indonesia), which is low compared to the recognition accuracy of 66% obtained, for example, in the cross-culture study by Scherer et al. (2001). Apparently, the emotion-related acoustic differences were too subtle to be detected in this analysis. As play-acted stimuli were no more acoustically distinct than the authentic ones, the authenticity of half of the stimuli is not an explanation for low emotionality. In contrast to other studies in which the speakers were asked to express the emotion in one sentence (Scherer et al., 2001; Laukka et al., 2005) or in one word (Leinonen et al., 1997), the speakers of this study performed long speech sequences without knowing which part would be used in the analysis. As Hammerschmidt and Jürgens (2007) noted, emotions are not equally encoded in every single-word, and it therefore seems plausible that the emotionality was spread over the complete sequence and did not become as evident in the single-word expressions. Additionally, it is plausible that the categorizations of the spontaneous expressions into the four emotion categories ("anger," "sadness," "fear," and "joy") might be artificial (Laukka et al., 2011). To complicate matters further, it is rarely the case that only one emotion is encoded in spontaneous speech (Greasley et al., 2000).

The fact that we were not able to detect emotionality at all deserves special attention. Due to this, we cannot exclude the possibility that the emotional expressivity is influenced by authenticity. An analysis of longer and more exaggerated stimuli is needed to shed light on the question whether the vocal expression of the different emotions is similar between the encoding conditions in every detail. Our analysis, in any case, suggests that such an effect would probably be very subtle and that the effect of the general encoding differences is much more dominant.

CONCLUSION

This study revealed that during the acting process a type of speech is used that differs from the one during spontaneous expressions. We demonstrated that play-acted expressions are not necessarily encoded in an exaggerated, stereotypical or more aroused fashion in comparison to naturally occurring expressions, as proposed before (Scherer, 2003). Instead, it appears that the acting process affects the vocal expression in a more general way. Therefore, caution should be exercised when using emotion portrayals by professional actors, as in combination with the study that concentrated on the listeners' perspective (Scheiner and Fischer, 2011) it emerged that encoding differences lead to an influence on the emotion perception in terms of play-acted stimuli. Future research should aim to uncover what the causes of these differences in encoding are and whether listeners make use of these acoustic cues to judge authenticity and emotion.

REFERENCES

Audibert, N., Aubergé, V., and Rilliard, A. (2008). "How we are not equally competent for discriminating acted from spontaneous expressive speech," in *Speech Prosody 2008*, Campinan.

Audibert, N., Aubergé, V., and Rilliard, A. (2010). "Prosodic correlates of acted vs. spontaneous discrimination of expressive speech: a pilot study," in *5th International Conference on Speech*, Chicago.

Bachorowski, J., and Owren, M. J. (1995). Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. *Psychol. Sci.* 6, 219–224.

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636.

Barkhuysen, P., Krahmer, E., and Swerts, M. (2007). "Cross-modal perception of emotional speech," in *International Congress of Phonetic Sciences*, Saarbrücken.

Boersma, P., and Weenink, D. (2009) *Praat: Doing Phonetics By Computer (Version 5.1.11) [Computer program]*. Available at: http://www.praat.org/ [Retrieved August 4, 2009].

Botinis, A., Granström, B., and Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Commun.* 33, 263–296.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* 18, 32–80.

Ekman, P., and Oster, H. (1979). Facial expressions of emotion. *Annu. Rev. Psychol.* 30, 527–554.

Ekman, P., and O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: the utility of voluntary and involuntary behavior in detecting deception. *Behav. Sci. Law* 24, 673–686.

Ekman, P., Sorenson, E. R., and Friesen, W. V. (1969). Pan-cultural elements

in facial displays of emotion. *Science* 164, 86–88.

Elfenbein, H. A., Foo, M. D., Mandal, M., Biswal, R., Eisenkraft, N., Lim, A., and Sharma, S. (2010). Individual differences in the accuracy of expressing and perceiving nonverbal cues: new data on an old question. *J. Res. Pers.* 44, 199–206.

Goudbeek, M., and Scherer, K. R. (2010). Beyond arousal: valence and potency/control cues in the vocal expression of emotion. *J. Acoust. Soc. Am.* 128, 1322–1336.

Greasley, P., Sherrard, C., and Waterman, M. (2000). Emotion in language and speech: methodological issues in naturalistic approaches. *Lang. Speech* 43, 355–375.

Hammerschmidt, K., and Jürgens, U. (2007). Acoustical correlates of affective prosody. *J. Voice* 21, 531–540.

Hanson, H. M., and Chuang, E. S. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *J. Acoust. Soc. Am.* 106, 1064–1077.

Jacewicz, E., Fox, R. A., O'Neill, C., and Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Lang. Var. Change* 21, 233–256.

Juslin, P. N., and Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion* 1, 381–412.

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814.

Kappas, A., and Polikova, N. (2008). "Judgments of the affective valence of spontaneous vocalizations: the influence of situational context," in *Emotions in the Human Voice*, Vol. 1, ed. K. Izdebski (San Diego, CA: Plural Publishing), 109–122.

Klein, J. (2010). Emotionstheater? Anmerkungen zum Spielgefühl. *Forum Mod. Theater* 25, 77–91.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J. Acoust. Soc. Am.* 59, 675–678.

Laukka, P., Audibert, N., and Aubergé, V. (2007). "Graded structure in vocal expression of emotion: what is meant by 'prototypical expression'?," in *1st International Workshop on Paralinguistic and Speech – Between Models and Data*, Saarbrücken.

Laukka, P., Juslin, P. N., and Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cogn. Emot.* 19, 633–653.

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., and Elenius, K. (2011). Expression of affect in spontaneous speech: acoustic correlates and automatic detection of irritation and resignation. *Comput. Speech Lang.* 25, 84–104.

Leinonen, L., Hiltunen, T., Linnankoski, I., and Laakso, M. L. (1997). Expression of emotional-motivational connotations with a one-word utterance. *J. Acoust. Soc. Am.* 102, 1853–1863.

Master, S., De Biase, N., Chiari, B. M., and Laukkanen, A. M. (2008). Acoustic and perceptual analyses of Brazilian male actors' and nonactors' voices: long-term average spectrum and the "Actor's Formant." *J. Voice* 22, 146–154.

Nawka, T., Anders, L. C., Cebulla, M., and Zurakowski, D. (1997). The speaker's formant in male voices. *J. Voice* 11, 422–428.

Owren, M. J. (2008). GSU Praat tools: scripts for modifying and analyzing sounds using Praat acoustics software. *Behav. Res. Methods* 40, 822–829.

Owren, M. J., and Bachorowski, J. A. (2007). "Measuring emotion-related vocal acoustics," in *Handbook of Emotion Elicitation and Assessement*, eds J. Coan and J. J. B. Allen (Oxford: Oxford University press), 239–266.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd Edn. Cambridge: Cambridge University Press.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *J. Acoust. Soc. Am.* 123, 1104–1113.

Scheiner, E., and Fischer, J. (2011). "Emotion expression – the evolutionary heritage in the human voice," in *Interdisciplinary Anthropology: The Continuing Evolution of Man*, eds W. Welsch, W. Singer, and A. Wunder (Heidelberg: Springer), 105–130.

Scherer, K. R. (1986). Vocal affect expression: a review and model for future research. *Psychol. Bull.* 99, 143–165.

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256.

Scherer, K. R., Banse, R., and Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross Cult. Psychol.* 32, 76–92.

Schrader, L., and Hammerschmidt, K. (1997). Computer-aided analysis of acoustic parameters in animal vocalisations: a multi-parametric approach. *Bioacoustics* 7, 247–265.

Sobin, C., and Alpert, M. (1999). Emotion in speech: the acoustic attributes of fear, anger, sadness and joy. *J. Psycholinguist. Res.* 28, 347–365.

Strasberg, L. (1987). *A Dream of Passion: The Development of the Method*. Boston: Little Brown.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *J. Acoust. Soc. Am.* 85, 1699–1707.

Velten, E. (1968). A laboratory task for induction of mood states. *Behav. Res. Ther.* 6, 473–482.

Waaramaa, T., Laukkanen, A. M., Airas, M., and Alku, P. (2010). Perception of emotional valences and activity levels from vowel segments of continuous speech. *J. Voice* 24, 30–38.

Westermann, R., Spies, K., Stahl, G., and Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: a meta-analysis. *Eur. J. Soc. Psychol.* 26, 557–580.

Williams, C. E., and Stevens, K. N. (1972). Emotions and speech: some acoustical correlates. *J. Acoust. Soc. Am.* 52, 1238–1250.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 February 2011; accepted: 19 July 2011; published online: 28 July 2011.
Citation: Jürgens R, Hammerschmidt K and Fischer J (2011) Authentic and play-acted vocal emotion expressions reveal acoustic differences. Front. Psychology 2:180. doi: 10.3389/fpsyg.2011.00180
This article was submitted to Frontiers in Emotion Science, a specialty of Frontiers in Psychology.

APPENDIX

APPENDIX I

Examples for transcripts

- Male spoken anger

Context

Two fighting dogs attacked 6-year-old V. in the schoolyard. He was bitten to death. Fighting dogs are a big problem in the area and people do not feel protected by the police. They are furious and are looking for a culprit. The anger is directed to the police. The people are shouting at a police officer, blaming him for being too late.

Man:

Original (German): Der Kiosk ruft vor Viertelstund an, "nach Viertelstund" kommt ihr erst, oder was?"

Translation: The kiosk called 15 minutes ago, you only come "after 15 minutes" or what?

- Female spoken anger

Context

Two freight trains crushed in the station of Bad Münder. One contained the toxic chemical Epichlorohydrin that leaked into the environment. The inhabitants are extremely angry about the poor flow of information. Nobody told them how dangerous the chemical really is and nobody seemed to think about the future effects the toxic substance in the ground could have on the people's health. One woman said very angrily:

Woman:

Original (German): Ham die kein schlechtes Gewissen sich hier hinzustellen? Zu sagen, wir kennen diese Substanz nicht? Die Kinder dürft ihr aber ruhig "auf die Wiese spielen lassen". Ist ja gar nicht schlimm.

Translation: Don't they feel any remorse to stand here? To say that they don't know the substance? But it is quite alright to let the children "play on the lawn". It's not that bad.

- Male spoken sadness

Context

In a school massacre, R. killed 16 people before killing himself. One of the victims was the spouse of E. One year after the gun rampage he reports on how he lives with the loss of his spouse. It is still difficult for him to accept her death. Besides his grief, he also felt a heavy pain about the helplessness of his friends.

Man:

Original (German): Und sie sitzen da plötzlich in Tränen "und alles lacht", oder ist dann plötzlich sprachlos, weil das dann passiert und sie das einfach nicht steuern können.

Translation: And you are sitting there in tears "and everybody laughs" or are speechless, because it just happened and you are not able to control it.

- Female spoken sadness

Context

The 73-year-old W. was attacked in his shop by two 16-year-old boys. He was robbed and stabbed to death. It is the date of the funeral. A weeping woman reports.

Woman:

Original (German): "Ich kenn den 43 Jahr." Und er war für uns alle ein Freund. Und ich finde das furchtbar, was da passiert ist.

Translation: "I have known him for 43 years". And he was a friend, for all of us. And I think what happened is dreadful.

- Male spoken joy

Context

The Fall of the Berlin Wall. A citizen of the German Democratic Republic reports excitedly and happily about the border crossing.

Man:

Original (German): Vorhin haben sie noch einzeln durchgelassen. Dann haben sie das Tor aufgemacht, "und jetzt konnten wir alle" so, wie wir waren, ohne vorzeigen, ohne alles, konnten wir gehen.

Translation: Previously they let the people pass individually. Then they opened the gate and "now we could all", as we were, without showing anything, without everything, we could go.

- Female spoken joy

Context

A married couple has won a new car in the lottery. They report what trick they used to get their ticket on top for the drawing.

Woman:

Original (German): Ja. "Mein Mann wollte schon immer im Anfang" der ganzen Sache die 5 Mark einzahlen. Ich schob das ja immer noch n Bischn hinaus. Eben n kleiner Schnack, ne.

Translation: Yes. "My husband already at the beginning" of the whole thing wanted to pay the 5 Mark. Well, I always delayed that a bit. Just a little joke, huh.

- Male spoken fear

Context

Eleven miners had been buried 10 days before and were to be rescued by an additional drill. They were asked to stock up on lamps and food, leave the area, and withdraw to a deeper cave. Otherwise, the miners were strongly at risk of being injured by falling rocks. But the men refused to go down to the deeper cave. They were deeply afraid to be trapped in the small cave by the rocks from the drill. They communicate their fear to the operation controllers.

Man:

Original (German): Die Halde ist viel zu kurz und viel zu kurz abgestützt. "Weil der Tunnelbau, den wir hier abgestützt haben, mit dieser Folie". Der würde den Tunnelbau unmöglich ab …äh also… höchstwahrscheinlich abfangen. Aber wer kann dafür garantieren.

Translation: The acclivity is too short and supported much too short. "Because the tunnel construction, which we have supported, with this screen". It would impossibly... er, well… very likely hold back the tunnel construction. But who can guarantee that.

- Female spoken fear

Context

The 100 year flood at the Oder threatens whole villages. The water is rising and an inhabitant of an especially low-lying house reports her fears.

Woman:

Original (German): Grade unser Haus liegt ziemlich tief. Also 1947 stand das Wasser da schon "bis zum Fensterkreuz". Und wenn das noch schlimmer werden sollte, schätz ich, dass das Haus bald gar nicht mehr zu sehen ist im Wasser. Ja, ich hab ganz dolle Angst

Translation: Especially our house lies pretty low. Well, 1947 the water was already up to the window crossbar. And if it should get worse, I guess, that the house won't be visible anymore in the water. Yes, I am very much afraid.

Speech sequences were partly shortened. Only words in quotation marks were used for the analysis.

APPENDIX II

Table A1 | Abbreviations and descriptions of the acoustic parameters with high factor loadings.

Parameter	Factor	Description
pf mean, max, min (Hz)	F1	Mean, maximum, and minimum of the frequencies with the highest amplitude across time segments (PF)
diff mean, max, min (Hz)	F1	Mean, maximum, and minimum differences between F_0 and PF
dfa1 mean, max, min (Hz)	F1	Mean, maximum, and minimum frequency, frequency at which the amplitude distribution reaches the first quartile across all time segments (distribution of frequency amplitudes = dfa)
dfa2 min (Hz)	F1	Minimum frequency at which the amplitude distribution reaches the second quartile
fp1 mean (HZ)	F1	Mean frequency of the first global frequency peak
pf total max, min (Hz)	F1	Frequency of the total maximum and the total minimum amplitude
f2 mean (Hz)	F2	Mean frequency of the second global frequency peak
Range mean, max, min (Hz)	F2	Difference between highest and lowest frequency within a segment, mean across time segments, maximum, minimum
dfa2 mean, max (Hz)	F2	Mean and maximum frequency at which the amplitude distribution reaches the second quartile
dfa3 mean, max, min (Hz)	F2	Mean, maximum, and minimum frequency at which the amplitude distribution reaches the second quartile
pf jump	F3	Maximum differences between successive PFs
pf trend mean, max	F3	Mean and maximum deviation between pf and linear trend
F_0 mean, max, min (Hz)	F4	Mean, maximum, and minimum fundamental frequency across tonal time segments
Tonality (%)	F5	Percentage of tonal time segments
BWF1 (Hz)	F6	Bandwidth of the first formant
amprat2	F7	Amplitude ratio between second harmonic and F_0
amprat3	F7	Amplitude ratio between third and first harmonic
HNR mean, max	F8	Differences between highest and lowest frequency within a segment, mean across all time segments, maximum (1 = no noise)
Jitter mean, max	F9	Mean and maximum cycle-to-cycle variations in the F_0, across all time segments
dfa1 max location	F10	Relative position of the maximum value of the first dfa1 (0 = beginning of the call and 1 = end of the call) [(1/duration) × location]
dfa2 max location	F10	Relative position of the maximum value of the second dfa1 [(1/duration) × location]
Shimmer mean, max	F11	Mean and maximum cycle-to-cycle variations in the amplitude, across all time segments
cs max location	F12	Relative position of the maximum value of the correlation coefficient of successive time segments [(1/duration) × location]
pf minimum location	F12	Relative position of the minimum value of the peak frequency [(1/duration) × location]
cs mean	F13	Mean correlation coefficient of successive time segments

Detailed descriptions were partly taken from Hammerschmidt and Jürgens (2007).