Emotion dimensions and formant position

Martijn Goudbeek[1,3]*, Jean Philippe Goldman*[2]*, Klaus R. Scherer*[3]

[1] Department of Communication and Information Sciences, Tilburg University, The Netherlands
[2] Language Technology Laboratory, University of Geneva, Switzerland
[3] Swiss Center for Affective Sciences, Switzerland

m.b.goudbeek@uvt.nl, goldman@lettres.unige.ch, klaus.scherer@unige.ch

Abstract

The influence of emotion on articulatory precision was investigated in a newly established corpus of acted emotional speech. The frequencies of the first and second formant of the vowels /i/, /u/, and /a/ was measured and shown to be significantly affected by emotion dimension. High arousal resulted in a higher mean F1 in all vowels, whereas positive valence resulted in higher mean values for F2. The dimension potency/control showed a pattern of effects that was consistent with a larger vocalic triangle for emotions high in potency/control. The results are interpreted in the context of Scherer's component process model.

Index Terms: Articulation, formants, emotion, nonverbal expression

1. Introduction

The emotional state of the speaker is reflected in his or her vocal utterances. Listeners know this and are able to recognize emotional states based on vocal cues alone [1]. The vocal information used to recognize affective states is traditionally divided in four groups: (1) temporal information, (2) pitch information, (3) intensity information, and (4) voice quality information. Voice quality is primarily associated with the spectral properties of the speech signal [2, 3]. In this paper we focus on formant position, a spectral property of the speech signal that reflects voice quality [3] as well as linguistic vowel identity [4].

While most research to date has investigated the effect of emotion on full utterances [2], this paper presents, for the first time, detailed analyses at the segmental (vowel) level of the effect of emotion dimension on vocal expression.

Speech is a signal that conveys verbal as well as nonverbal information. In emotional speech important carriers of the linguistic message need to be preserved for the verbal message to get across, while properties of the speech signal that are less crucial for the phonetic identity of speech segments are more likely affected by emotion. Here, we take a first step in the study of this interaction between verbal and nonverbal information in the speech signal by investigating the effect of emotion dimensions on formant placement in individual vowels.

Formants are those regions in the spectrogram where the amplitude of the acoustic energy is high. They reflect the natural resonance frequencies of the vocal tract. These natural resonances are not fixed, but can be changed by altering the shape of the vocal tract. The resonances of the vocal tract can be changed, for example, by changing the position of the tongue or the jaw [4]. Formants are numbered from the lowest frequency upward (F1, F2, F3, etc.). A spectrogram can contain as many as ten identifiable formants, but most analyses do not go beyond the first five. Here we report on the most important formants for vowel identification, the first and second formant (F1 and F2). These two formants can be used to uniquely determine vowel identity. Table 1 lists the formant frequencies of the cardinal vowels for French [5].

Together the cardinal vowels /a/, /i/, and /u/ constitute the vowel space that contains the articulatory space for all the possible vowels. The size of the vowel space is not fixed. For instance, it is modulated in infant directed speech [6]. Mothers talking to infants not only raise their pitch, but also stretch the distance between the vowels /a/. /i/. and /u/. By exaggerating the distance between the vowels mothers make it easier for infants to separate the vowels into contrasting categories and they simultaneously highlight the parameters (F1 and F2) that differentiate these vowels [6].

Table 1. *The three cardinal vowels /a/, /i/, and /u/, and their mean formant frequencies for French.*

Vowel	F1 (Hz)	F2 (Hz)
a	756	1391
i	277	2321
u	300	1863

Infant directed speech serves to teach infants about the sounds of their language, but it is also serves to communicate (positive) affect between mother and child [7]. An experiment that had mothers talk to infants as well as to pets showed that while both types of speech were rated as high in affect, only infant directed speech had an increased vowel triangle [7]. The authors concluded that an increased vocalic area was a didactic device rather than a nonverbal affective cue.

However, in [8] we measured the size of the vocalic area of twelve emotions that differed parametrically in valence, potency and arousal. The analysis showed that different emotions *do* differ in the size of their vocalic triangle. In particular, emotions that score high on the dimension potency/control have a larger vocalic triangle than emotions that are low on potency/control [9, 10]. A study with a different, smaller, corpus also found differential effects of emotion on the size of the vocalic area [11]. In this contribution, we investigate the effect of emotion dimension on the placement of F1 and F2.

The emotions present in our corpus can be divided along three dimensions: valence, arousal and potency/control. (see Table 3) Valence and arousal are often studied dimensions in emotion research and part of almost any theory of emotion [12, 13]. Valence refers to the intrinsic attractiveness (positive valence) or aversiveness (negative valence) of an event, object or episode leading to an emotion [13,14].

Arousal represents the degree of alertness, excitement or engagement with the object(s) of emotion [13]. Arousal is often found to have a large effect on vocal expression. [2], often obfuscating effects of valence or potency/control. Previous investigations have tried to statistically control for arousal by regression on intensity measures [2], but the current corpus [15] enables the parametric analysis of arousal level.

Potency/control also has a long history in emotion research and refers to the individual's sense of power or

control over the events [13, 16]. Its role as an important third factor has has recently been highlighted in cross-linguistic work on emotion terms [9].

The component process model [1,10] is the only model of emotion that predicts testable effects of emotion on vocal expression. In the CPM emotional episodes are the result of a sequence of appraisal checks for novelty, pleasantness, goal significance and conduciveness, coping potential and norm compatibility. All psychological and physiological effects accompanying an emotional episode are considered to be direct effects of these appraisal checks.

Table 2. *The predictions of the CPM for the effect of the appraisal checks intrinsic pleasantness (valence) and coping potential (potency/control) on formant position.*

Intrinsic pleasantness check (Valence)	
Pleasant:	F1 falling
Unpleasant:	F1 rising, F2 falling
Coping potential check (Potency/control)	
High :	Pronounced formant differences
Low	Formant frequencies neutral

With pleasant emotions (positive valence), the CPM predicts a drop in F1, whereas with unpleasant emotions F1 is predicted to rise. In addition, F2 is predicted to go down for unpleasant emotions. Based on this, we predict a lower F1 and higher F2 for all vowels for emotions with positive valence compared to emotions with negative valence.

In situations where the speaker perceives control, the CPM predicts pronounced differences between the formant frequencies. In contrast, when the speaker does not perceive control, the formant frequencies will tend to their neutral settings. Based these predictions and previous work [8], the effect of potency/control is predicted to depend *both* on vowel and formant, so that the changes in formant frequencies will results in a larger vocalic area for emotions high in potency/control. Specifically, the F1 of /a/ is predicted to be higher for emotions high in potency/control, while the F1 of /u/ and /i/ is predicted to be lower. Similarly, the F2 of /a/ and /u/ is predicted to be lower and higher for /i/ for emotions high in potency/control.

There is no appraisal check in the CPM that maps directly onto the arousal dimension. Given that arousal reflects physiological processes that often result in higher values for fundamental frequency and intensity [2], high aroused emotions are predicted to have elevated values for F1 and F2 for all vowels.

2. Method

2.1. Corpus

We used a subset of the Geneva Multimodal Emotion Portrayals [15] to investigate the effect of emotion dimension on formant placement. The subset consisted of 12 emotions expressed by ten actors (five males and five females) from diverse age groups. The actors expressed the emotions in two nonsense utterances: [ne kali bam sud molɛn] and [kun se mina lud belam].

Table 3 shows the division of the emotions present in the corpus into the three dimensions of emotion: arousal, valence, and potency control [1, 8, 10]. We will exploit this property of the corpus by investigating the effect of emotion dimension instead of the effect of each individual emotion.

The emotions were portrayed by the actors in close interaction with a professional stage director, informed by the principles put forward by Stanislawski [2]. This way, the rendition of the emotion was as naturalistic as possible.

Table 3. *The twelve emotions present in the analyzed corpus in a valence times arousal grid. Emotions high on the dimension potency/control are in boldface.*

	Valence	
	Positive	Negative
	elation (joy)	**anger (ang)**
	amusement (amu)	panic fear (fea)
	pride (pri)	despair (des)
	pleasure (ple)	**irritation (irr)**
	relief (rel)	anxiety(anx)
	interest (int)	sadness (sad)

2.2. Portrayal selection

Several utterances were recorded for each actor. A rating study on 1260 preselected portrayals was conducted to select the best expressions of each emotion for each actor. Criteria for inclusion in the present corpus depended on recognition accuracy of the intended emotion and the believability of the portrayal. Because raters could select more than one emotion per portrayal in the rating study, we calculated a recognition-index that took into account the recognition accuracy for the target emotion, the recognition accuracy for the second highest category and the total amount of ratings. For each emotion, one portrayal by each actor was included (resulting in 10 actors x 12 emotions = 120 portrayals).

2.3. Segmentation and formant extraction

All portrayals were manually segmented at the phoneme level by a linguist collaborating with the authors on the annotation of the corpus. After segmentation, formants were extracted for the vowels /a/, /i/, and /u/ using PRAAT speech analysis software with standard settings [17]. These include the maximum number of formants tracked (five), the maximum frequency of the highest formant (set to 5000 for male and 5500 for female speakers), the time step between two consecutive analysis frames (0.01 seconds), the effective duration of the analysis window (0.025 seconds) and the amount of pre-emphasis (50 Hz). These settings generally resulted in acceptable results. Nevertheless, we manually corrected the erroneous formant estimations, notably confusing F3 for /u/ with F2, by replacing them with the value observed in the spectrogram.

3. Results

The extracted values for F1 and F2 were analyzed separately for the vowels /a/, /i/, and /u/. For each dependent variable (F1 and F2) we performed a within-subjects analysis of variance with three within-subject factors: valence (positive versus negative), arousal (high versus low) and potency/control (high versus low).

The figures 1 to 6 display the mean values of F1 and F2 for the three vowels and the three dimensions. In all six analyses of variance, there was only one significant two-way interaction and no significant three-way interactions. The only significant two-way interaction was the one between valence and arousal for F1 of /u/ and reflects the larger difference in F1 values between high and low levels of valence at low levels of arousal. None of the directions of the main effects were affected by this interaction. Hence, only main effects are reported below.

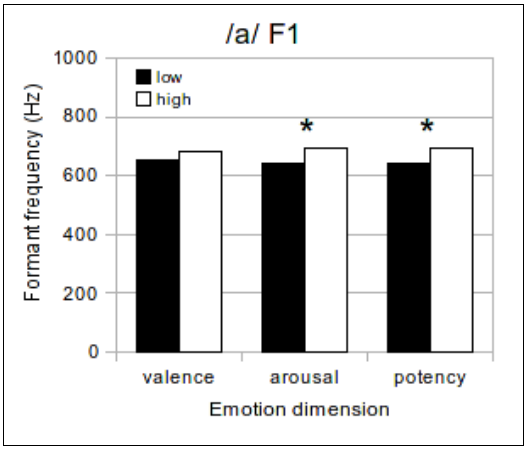

Figure 1: *Extracted values for F1 for the vowel /a/ and the three emotion dimensions. The asterisks indicate significant differences at the 0.05 level.*

Figure 1 displays the values for F1 for the vowel /a/ for the emotion dimensions valence, arousal and potency/control (titled potency). Emotions that are positive in valence, high in arousal or high in potency/control have higher mean values for F1. This difference is significant for high and low aroused emotions ($F [1,9] = 7.57$, $p < 0.022$) and for emotions that differ in their level of potency/control ($F [1,9] = 10.36$, $p < 0.011$).

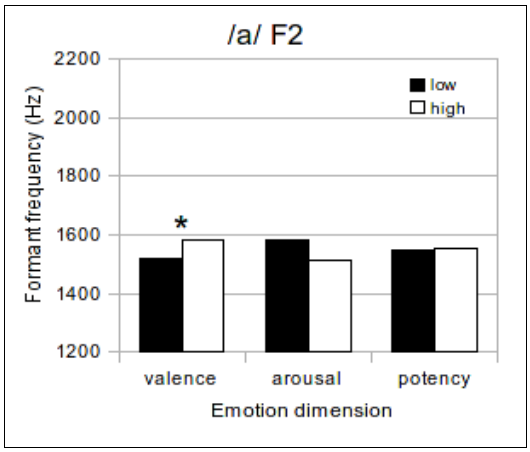

Figure 2: *Extracted values for F2 for the vowel /a/ and the three emotion dimensions. The asterisks indicate significant differences at the 0.05 level.*

Figure 2 displays the values for F2. for the vowel /a/. It is important to note that the scales of Figure 1 and Figure 2 differ in their range (due to the different target frequencies for F1 and F2) and that the starting point of the y-intercept is not zero but 1200 for Figure 2. Here, positive valence leads to a significantly higher value for F2 ($F [1,9] = 4.92$, $p < 0.05$), whereas high levels of arousal lead to a lowering of F2 ($F [1,9] = 4.68$, $p < 0.05$). Potency control does not influence the value of the second formant of /a/.

The mean values for F1 for each emotion dimension for the vowel /i/ are displayed in Figure 3. The effects of emotion dimension for the first formant of /i/ are similar to those for the first formant of /a/. High levels for arousal ($F [1,9] = 15.75$, $p < 0.003$) and potency/control ($F [1,9] = 8.84$, $p < 0.016$). result in significantly elevated values for F1.

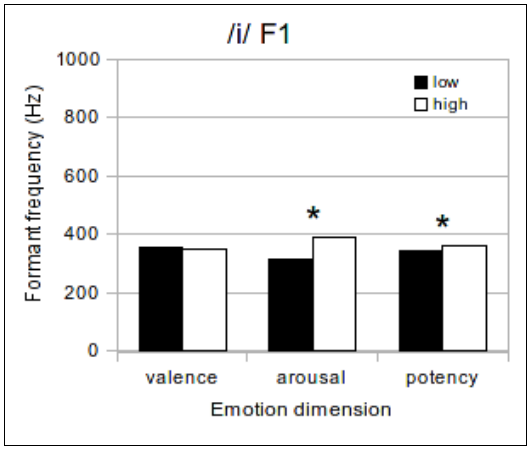

Figure 3: *Extracted values for F1 for the vowel /i/ and the three emotion dimensions. The asterisks indicate significant differences at the 0.05 level.*

Figure 4 shows the mean values for F2 per emotion dimension for the vowel /i/. There were no significant effects of emotion dimension on the value of the second formant. However, the difference between positive and negative valence did approach significance ($F [1,9] = 3.24$, $p < 0.110$).

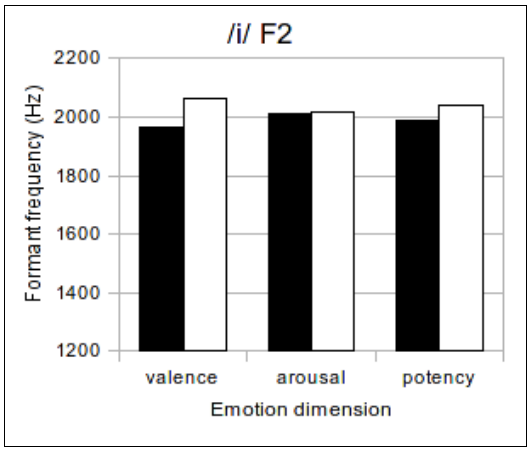

Figure 4: *Extracted values for F2 for the vowel /i/ and the three emotion dimensions.*

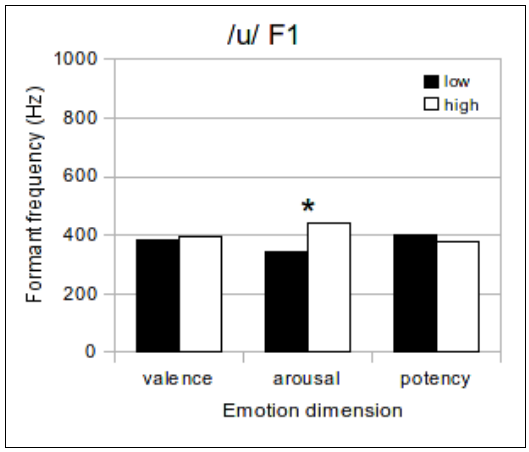

Figure 5: *Extracted values for F1 for the vowel /u/ and the three emotion dimensions. The asterisk indicates the significant difference at the 0.05 level.*

Figure 5 shows the mean F1 values for the vowel /u/ for each emotion dimension. The pattern of effects is slightly different from that observed in the values of F1 for /i/ and /a/. The significant effect of the dimension arousal is in the same direction as for /a/ and /i/: high aroused emotions have a significantly higher mean F1 than low aroused emotions ($F[1,9] = 18.38$, $p < 0.002$).

The mean values for F2 of /u/ per emotion dimension are displayed in Figure 6. The mean value of F2 differs significantly for valence ($F[1,9] = 9.47$, $p < 0.013$) and potency/control ($F[1,9] = 9.22$, $p < 0.014$) on the value of F2. Positive valence is reflected in a higher value for F2, whereas high potency is reflected in a lower value for F2.

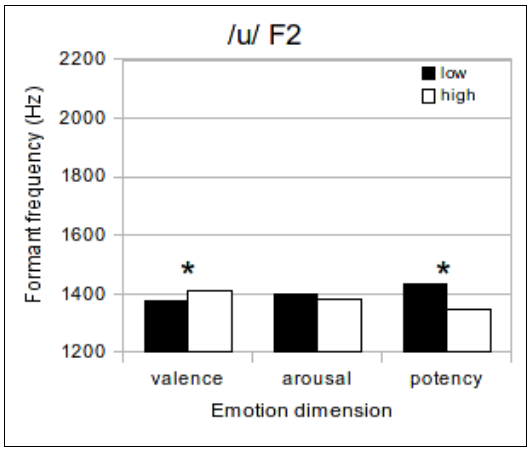

Figure 6: *Extracted values for F2 for the vowel /u/ and the three emotion dimensions. The asterisks indicate significant differences at the 0.05 level.*

4. Discussion

This work is a first investigation into the effects of emotion dimensions on specific phonetic segments (but see 18 and 11 for work on emotional states and phonetic segments). Since the parameters investigated (F1 and F2) have an important function in determining vowel identity, the research highlights the interplay between phonetic and affective factors in speech production.

The results show that emotion has a sizable influence on formant positioning. The mean of F1 of all three vowels /a/. /i/, and /u/ is higher for high aroused emotions and lower for low aroused emotions. In addition, high aroused emotions have a significantly lower F2 for /a/. These effects confirm the important role of arousal in the vocal expression of emotion.

Formant positioning is also significantly affected by the valence dimension. Positive emotions have a higher mean F2 than negative emotions. This holds for all vowels, although the effect is only significant in /a/ and /u/.

The potency control dimension shows a mixed picture of effect with high potency/control resulting in higher values for F1 in /a/ and /i/, but lower values for mean F2 in /u/. With the exception of the higher F1 for /i/, all these changes in formant position contribute to an increased vocalic triangle. Especially the large drop in F2 for /u/ results in a larger vocalic area.

These results by and large confirm the predictions in that the coping potential check does lead to more widely spaced harmonics. For the pleasantness check (valence) the results counter the predictions of the CPM for F1: positive emotions results in a raised F1. However, the the results are in line for F2: negative emotions results in lower values for F2. The elevated values for F1 for high aroused emotions also seem to be in line with the CPM.

In sum, these results complement previous work that showed potency/control to be related to the area of the vocalic triangle. The emotion dimensions valence and arousal are reflected in formant positioning as well, but do not result in a larger vocalic area. Instead, positive valence and high arousal result in heightened values of F2 and F1 respectively.

5. Acknowledgements

We thank Tanja Bänziger for the work done on construction and rating of the corpus, Marcello Mortillaro for the selections based on the lay ratings, Hannes Pirker for the phonetic segmentation of the corpus and Emiel Krahmer for careful reading of and commenting on the manuscript. This research has been funded by Vici grant 277-70-007 from The Dutch Scientific Research Council and by SNFS grant 101411-100367 by the SNFS National Competence Center in Affective Sciences.

6. References

[1] Scherer, K.R., "Vocal affect expression: A review and a model for future research", Psychological Bulletin, 1986, 99, 143-165.

[2] Banse, R., and Scherer, K.R., "Acoustic profiles in vocal emotion expression." J Pers Soc Psychol., 1996, 3, 614-636.

[3] Gobl, C.& Ní Chasaide, A. N., "The role of the voice quality in communicating emotions, mood and attitude" Speech Comm., 2003, 40, 189-212.

[4] Stevens, K.N. "Scientific substrates of speech production", In F.D. Minifie (Ed.), Introduction to Communication Sciences and Disorders, Singular, San Diego, CA, 1994, 399- 437.

[5] Dowd, A., Smith, J., and Wolfe, J., "Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time." Lang Speech, 1996, 41, 1-20.

[6] Kuhl, P., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.K., Stolyarova, E.I., Sundberg, U., Lacerda, F., "Cross-language analysis of phonetic units in language adressed to infants." Science, 1997, 5326, 684–686.

[7] Burham, D., Kitamura, C., and Vollmer-Conna, U., "What's new, pussycat? On talking to baby's and animals." Science, 2002, 296, 1435.

[8] Goudbeek, M., Goldman, J.P, and Scherer, K.R. "Emotions and articulatory precision." In Proceedings of Interspeech 2008 incorporating SST 2008, 317, Brisbane: ISCA.

[9] Fontaine, J. R., Scherer, K. R., Roesch, E. B., and Ellsworth, P. "The world of emotion is not two-dimensional." Psych Science, 2007, 13, 1050-1057

[10] Scherer, K.R. "Vocal communication of emotion: A review of research paradigms." Speech Comm., 2003, 40, 227-256.

[11] Beller, G., Obin, N., and Rodet, X., "Articulation degree as a prosodic dimension of expressive speech". In Proceedings of the fourth conference on Speech Prosody, 681, Campinas, Brazil, May 6-9. 2008.

[12] Russell, J.A., "A circumplex model of affect." J Pers Soc Psychol., 1980, 39, 1161–1178.

[13] Wundt, W., "Grundriss der psychologie. Achte auflage." [Outlines of psychology]. Leipzig, Germany: Engelmann, 1909.

[14] Frijda, N.H. "The Emotions." Cambridge (UK): Cambridge University Press, 1986.

[15] Bänziger, T., and Scherer, K., "Using actor portrayals to systematically study multimodal emotion expression: the gemep corpus.", In A. Pavia, R. Prada, R.W. Picard (Eds.), *Affective Computing and Intelligent Interaction,* Second International Conference, ACII, 2007, 476-487, Lisbon, Portugal.

[16] Osgood, C.E., May, W.H., and Miron, M.S., "Cross-Cultural Universals in Affective Meaning." University of Illinois Press, Urbana, IL, 1975.

[17] Boersma, P., and Weenink, D., "Praat: doing phonetics by computer." (Version 5.0.07) [Computer program], 2008. Retrieved February 1, 2008, from http://www.praat.org/.

[18] Lee, C.M., Yilderim, S., Bulut, Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., and Narayanan, S. "Emotion recognition based on phoneme classes." Proceedings of ICSLP, Jeju Island, Korea, October 2004.