# Validating a Multilingual and Multimodal Affective Database

Juan Miguel López[1], Idoia Cearreta[1], Inmaculada Fajardo[2], and Nestor Garay[1]

[1] Laboratory of Human-Computer Interaction for Special Needs (LHCISN).
Computer Science Faculty. University of the Basque Country
Manuel Lardizabal 1; Donostia - San Sebastian
[2] Cognitive Ergonomics Group
Department of Experimental Psychology. University of Granada
Cartuja Campus; Granada
juanmi@si.ehu.es, icearreta001@ikasle.ehu.es, ifajardo@ugr.es,
nestor.garay@ehu.es

**Abstract.** This paper summarizes the process of validating RekEmozio, a multilingual (Spanish and Basque) and multimodal (audio and video) affective database. Fifty-seven participants validated a sample of 2,618 videos of facial expressions and 102 utterances in the database. The results replicated previous findings of no significant differences in recognition rates among emotions. This validation has allowed having the audio and video material in the database classified in terms of the emotional category expressed. This normative data has proven to be useful for both training affective recognizers and synthesizers and carrying out empirical studies on emotions by psychologists.

**Keywords:** Affective computing, affective resources, user validation, multilingual and multimodal resources, semantics.

## 1 Introduction

Human beings are eminently emotional, as their social interaction is based on the ability to communicate their emotions and perceive the emotional states of others [1]. Affective computing, a discipline that develops devices for detecting and responding to users' emotions, and affective mediation, computer-based technology which enables the communication between two or more people, displaying their emotional states [2, 3], are growing areas of research [4].

Affective mediation tries to minimize the filtering of affective information carried out by communication devices, due to the fact they are usually devoted to the transmission of verbal information and therefore, miss nonverbal information [5]. Applications of mediated communication can be textual telecommunication technologies such as affective electronic mail, affective chats, etc.

In the development of affective applications, affective resources, such as affective stimuli databases, provide a good opportunity for training such applications, either for affective synthesis or for affective recognizers based on classification via artificial neural networks, Hidden Markov Models, genetic algorithms, or similar techniques

N. Aykin (Ed.): Usability and Internationalization, Part II, HCII 2007, LNCS 4560, pp. 422–431, 2007.

(e.g., [6, 7]). As seen in [8], there is a great amount of effort devoted to the development of affective databases. Affective databases usually record information by means of images, sounds, speech, psychophysiological values, etc. One of the main risks with affective databases is not having correctly labeled information. Therefore, they should be validated by human subjects in order to ensure that the stimuli adequately express the affects they are supposed to.

In this paper, the validation of the multilingual (Spanish and Basque) and multimodal (utterances and facial expressions videos) RekEmozio affective database is presented. In the following sections, affective models and related work are briefly revised. Next, RekEmozio database characteristics, validation process and main results are presented. Finally, a number of conclusions are outlined and future work is proposed.

## 2 Related Work

### 2.1 Models of Emotions

There are many possible ways where emotional parameters can be registered, codified or interpreted by computers. Models of emotions proposed by cognitive psychology could be a useful starting point. Generally speaking, models of emotions can be classified into two main groups: categorical and dimensional emotional models.

Categorical models of emotions have been more frequently used in affective computing (see [8] for a revision). In the research of emotions, different category groups related to emotions have been suggested. For example, authors such as [9] think that there are six basic emotions and that they are universal and shared by all humans, from which the rest of affective reactions are derived. These emotions, also called "Big-Six" emotions, are *anger, joy, sadness, disgust, fear* and *surprise*.

Dimensional approach to emotion has been advocated by a number of theorists, such as [10, 11]. Emotion dimensions are a simplified description of basic properties of emotional states [12]. The most frequent dimensions found in the literature are *Valence*, *Activation*, and *Control*. Therefore, a stimulus can be classified according to these 3 dimensions, for instance, it can be said that a concrete utterance has high valence, low activation and high control.

### 2.2 Affective Databases

Cowie and colleagues carried out a wide review of existing affective databases [8] which are described according to diverse features such as naturalness (e.g., emotion elicitation method) and scope (e.g., material: audio, video, mix; language).

Regarding material, there are databases of speech, sounds, text, faces or video scenes. With respect to speech, most references found in literature are related to English, while other languages have less resources developed, especially the ones with relatively a low number of speakers. This is the case of Basque. To our knowledge, the first affective database in Basque is the one presented by [13]. In Spanish, the work of [14] stands out.

Our understanding is that there is no validated database in Basque and Spanish which includes multimodal material (audio and video). Consequently, this type of

database is essential for research in affective recognition and production, a database called RekEmozio, which includes these features, is described in the next section.

# 3 RekEmozio Database

## 3.1 Database Description

The RekEmozio database was created with the aim of serving as an information repository for performing research on user emotion. Members of different work groups involved in research projects related to RekEmozio performed several processes for extracting speech and video features such as frequency, volume, etc. This information is described in [15]. The characteristics of the RekEmozio database are described in Table 1 [16].

**Table 1.** Summary of RekEmozio database features

| Scope | | | Naturalness | | | | Context |
|---|---|---|---|---|---|---|---|
| **Language** | **Description given of emotions** | **Number of actors/ actresses** | **Emotion elicitation methods** | **Semantically meaningful content** | **Material** | **Same text per emotion** | **Mode** |
| Spanish | sadness, fear, joy, anger, surprise, disgust; neutral | 10 (5/5) | Contextualized acting | Combined | 2,618 audio stimuli and 102 video stimuli | Non-semantically meaningful texts | Audio-Visual |
| Basque | | 7 (4/3) | | | | | |

As shown in Table 1, the RekEmozio database was created using recordings carried out by skilled bilingual actors and actresses. They received financial support for their cooperation. They were asked to read a set of words and sentences (both semantically and non-semantically relevant) trying to express emotional categories by means of voice intonation. The emotional categories considered are the classical "Big-Six" plus the neutral one, as shown in Table 1 ("Description given of emotion" column). In addition, they were asked to express facial expressions related to these emotional categories.

Regarding spoken material, the paragraphs and sentences used were constructed by using a group of words extracted from an affective dictionary in Spanish (1,987 words dictionary with nouns, adjectives, verbs and interjections). This emotional dictionary is built on top of words contained in the database of [17].

Semantically meaningful paragraphs and sentences were built from this group of words. Moreover, non-semantically meaningful words with the "neutral" label were used. For Basque sentence creation, sentences from Spanish were translated.

# 4 RekEmozio Database Validation

The procedure for performing the normative study to obtain affective values from the given audio-visual material is described next.

### 4.1 Method

#### 4.1.1 Participants

Fifty-seven volunteers participated in the validation, 36 men (average age of 26.25, sd=9.7; age range=17-56) and 21 women (average age of 27.5; sd=10.7; age range=18-52). The mother tongue of 31 participants was Spanish and the mother tongue of the remaining 26 participants was Basque. They received financial support for their cooperation.

#### 4.1.2 Material and Tools

A set of 2,720 stimuli were obtained from RekEmozio database, from which 2,618 were oral expressions (words, sentences and paragraphs) and 102 were videos with facial expressions. In order to ask subjects to validate affectively the stimuli, subjects were asked to select an emotional label for each stimulus (categorical test). Finally, in order to automate data recovery and facilitate the analysis of collected data, a tool called Eweb [18] was used.

*4.1.2.1 Categorical Test.* For the RekEmozio database validation, categorical measures were used, as the recordings within the database itself were performed by actors and actresses trying to express the above mentioned seven categorical emotions. Thus, when validating the database, human subjects were asked to indicate what emotion they thought the actors and actresses were attempting to express in each different database recording or stimulus.

*4.1.2.2 Instruments.* In order to automate data recovery and facilitate the analysis of collected data, Eweb [18, 19], a tool for designing and implementing controlled experiments in Human-Computer Interaction (HCI) environments, was used.

#### 4.1.3 Design

A mixed multifactorial design was followed. The *Language* (Spanish, Basque) variable and *Actor* variable (10 or 7 levels, depending on the number of actors per language) were manipulated between-groups, while *Emotion* (joy, sadness, anger, disgust, surprise, fear, neutral) and *Media* (audio, video) were manipulated within-subject. In the case of audio material, according to RekEmozio database features, two more variables were manipulated: *Text Length* (word, sentence, paragraph) and *Semantics* (semantically meaningful, non-semantically meaningful). Each subject validated 160 stimuli (154 oral expressions and 6 videos) corresponding to one single actor.

#### 4.1.4 Procedure

The participants used the interface provided by Eweb to perform their validation. First, participants received general and specific instructions for performing the experiment and they had to fulfil a demographic questionnaire. Afterwards, they began the session itself. Each participant performed the validation for one language, thus, they received the instruction about the language in which they had to perform.

The session was divided into two blocks (audio and video) and Eweb randomly presented each one to the participants. They had to perform several trial sessions for each block (three for oral and three for facial). They later performed the experimental session for the 154 stimuli in audio (where the audio block was selected) and 6 in video (otherwise). Participants had to complete a questionnaire for each stimulus, by selecting the emotional category. Each stimuli was heard/seen only once by the participants. They could only select one value for the category in the questionnaire. When participants finished with their first block, Eweb assigned them a second. The validation session finished after participants had completed both audio and video blocks. The procedure was the same for each language.

### 4.2 Results

#### 4.2.1 Emotion Recognition in Vocal Expression

First of all, the data in the categorical test was analyzed. Recognition accuracy percentages for the different types of utterances (depending on the language) are presented in Table 2. Replicating previous data [20], Fear and Disgust obtained the lowest success percentages while Neutral, Joy, Sadness and Anger obtained the highest ones.

With the aim of contrasting whether the differences between emotions and languages were significant, a multifactorial ANOVA was performed with *Emotion*, *Text Length* and *Semantic* as within-subjects variables and *Language* as between-group variable. The percentage of recognition in the categorical test was introduced as dependent variable. The main effects of *Emotion*, $F(6,330)=34.11$; $Mse=0.13$; $p<0.000$ and *Text Length*, $F(2,110)=44.37$; $Mse=0.05$; $p<0.000$, were significant. In the case of the *Emotion* variable, the highest percentage of recognition was for Sadness (76%), followed by Anger (73%), Joy (73%), Surprise (63%), Fear (51%) and Disgust (51%). On the other hand, subjects obtained a higher percentage of emotion recognition with Sentences (M= 70%) and Texts (M= 71%) than with Words (M= 61%), $F(1,55)=67.6$; $Mse= 0.07$; $p<0.000$. Finally, the main effect of *Semantic* was also significant, $F(1,55)=175.32$; $Mse= 0.1$; $p<0.000$.

**Table 2.** Recognition Accuracy Percentages for Utterances in function of language and emotions

| Language | Sadness | Fear | Joy | Anger | Surprise | Disgust | Neutral |
|---|---|---|---|---|---|---|---|
| **Spanish** | 75% | 51% | 78% | 71% | 66% | 52% | 80% |
| **Basque** | 77% | 52% | 68% | 74% | 59% | 51% | 77% |

Semantically meaningful texts obtained higher scores than non-semantically meaningful ones (77% and 58% of recognition respectively). There was neither significant main effect of *Language* nor interaction of this variable with others. Recognition percentages in function of *Emotion*, *Language*, *Semantic* and *Text Length* can be seen in Table 3 and 4.

**Table 3.** Recognition Accuracy Percentages for Spanish Utterances depending on language, emotions, semantic and text length

| Semantic | Text Length | Sadness | Fear | Joy | Anger | Surprise | Disgust | Neutral |
|---|---|---|---|---|---|---|---|---|
| Non-semantically Meaningful | Word | 68% | 38% | 44% | 58% | 63% | 39% | 76% |
| | Sentence | 72% | 33% | 74% | 78% | 49% | 25% | 74% |
| | Paragraph | 79% | 42% | 78% | 79% | 40% | 24% | 71% |
| Semantically Meaningful | Word | 80% | 53% | 63% | 67% | 63% | 52% | 77% |
| | Sentence | 83% | 75% | 81% | 90% | 68% | 93% | 78% |
| | Paragraph | 88% | 78% | 89% | 92% | 65% | 81% | 88% |

**Table 4.** Recognition Accuracy Percentages for Basque Utterances depending on language, emotions, semantic and text length

| Semantic | Text Length | Sadness | Fear | Joy | Anger | Surprise | Disgust | Neutral |
|---|---|---|---|---|---|---|---|---|
| Non-semantically Meaningful | Word | 57% | 41% | 71% | 53% | 64% | 38% | 79% |
| | Sentence | 76% | 38% | 72% | 68% | 62% | 23% | 70% |
| | Paragraph | 75% | 47% | 79% | 75% | 42% | 24% | 86% |
| Semantically Meaningful | Word | 75% | 52% | 75% | 65% | 71% | 62% | 76% |
| | Sentence | 93% | 68% | 92% | 91% | 83% | 89% | 73% |
| | Paragraph | 84% | 66% | 85% | 87% | 74% | 79% | 97% |

The 2-way interaction between *Emotion* and *Semantic* was significant, $F(6,330)=30.23$; $Mse=0.05$; $p<0.000$. The percentages of recognition for Fear (40%) and Disgust (20%) were especially lower in the case of non-semantically meaningful.

### 4.2.2 Emotion Recognition in Facial Expression

In the categorical test, the general accuracy percentage for facial expressions was 90% (sd 18.62). Table 5 shows the recognition accuracy percentage for each emotional category.

Fear and Disgust (as in the case of utterances) and Sadness were the emotions with the lowest recognition rates whereas Joy, Anger and Surprise obtained the highest recognition rates. In order to contrast whether the differences between emotions were significant, an ANOVA of repeated measures was performed with the six levels of *Emotion*. The difference between Sadness, Fear and Disgust on the one hand and Joy, Anger and Surprise on the other hand, was significant, $F(1, 56)=18.7$; $Mse=0.08$; $p<0.001$. Neutral results are not mentioned as neutral face is used as a reference in this study.

**Table 5.** Recognition Accuracy Percentages and Standard Deviations of facial expressions videos for each emotional category

| | Sadness | Fear | Joy | Anger | Surprise | Disgust |
|---|---|---|---|---|---|---|
| **Mean** | 81% | 81% | 98% | 96% | 96% | 89% |
| **SD** | 40% | 40% | 13% | 19% | 19% | 31% |

### 4.3 Discussion

The main goal of this study was to validate the audiovisual and multilingual emotional database *RekEmozio*. The data of the categorical test allows us to conclude that both audio and visual stimuli are valid to express the intended emotion as the recognition accuracy percentage is over 50% (78% in the case of audio and 90% in the case of video).

In the case of utterances, the differences in recognition accuracy between emotions found by other authors (e.g., [20]) were replicated, that is, Joy, Sadness and Anger are the type of utterances with the highest percentages of recognition while Fear and Disgust obtained the lowest recognition percentages. These relatively stable differences are supporting the hypothesis of the role of specific vocal parameters in the communication of different emotions [21]. In addition, the fact that Anger has the highest recognition percentages in vocal expression seems to agree with the results of [22]. The authors suggest an evolutionary explanation for these results: emotions that express danger, such as Anger and Fear, must be perceived at long distances with the aim of being perceived accurately by members of the group or even by the enemies. In order to do so, voice would be the most effective way, while facial expressions would be more effective for emotions which must be transmitted in short distances. This would explain why Anger (danger related) in vocal expressions is better recognized than other emotions not related to danger such as Surprise or Disgust. It would also explain why Joy and Surprise (which are supposed to be transmitted in short distances) obtained higher recognition rates in the case of facial expressions (98% and 96% respectively) than in the case of vocal expressions (75% and 62% respectively). However, there are two results which are incoherent with the evolutionary hypothesis: 1) Fear (danger related) presents one of the lowest recognition rates of vocal expressions; 2) Anger recognition rates are also higher in the case of facial expressions than in the case of vocal expressions. It would suggest that actors simply interpreted facial expressions better than vocal ones.

Another important finding of this study was the significant effect of semantic in the case of utterances. The semantically meaningful texts where better recognized than the non-semantically meaningful ones. Nevertheless, non-semantically meaningful texts, obtained an accuracy percentage of 58%, which is closer to the values obtained in earlier work (e.g., [21]). This means that, although the semantic content is an important contributor, the vocal or prosodic cues contribute much more effectively to the decoding of certain emotions. However, the recognition percentage for Fear (40%) and Disgust (20%) were especially low in the case of non-semantically meaningful utterances. According to [21], the low rates could be due to the diversity of modalities involved in the expression of these emotions (e.g., in the case of Disgust: nasal, visual, oral, semantic, etc.).

Finally, another goal of this study was to compare emotion recognition patterns in different languages, in this case Spanish and Basque. No relevant differences were found between languages. This agrees with the findings of other authors such as [23, 24, 25], who found that there is little difference in emotion detection between subjects coming from different linguistic and cultural environments, but state that recognition success rate in users is far from perfect. Therefore, in spite of being very different languages regarding to their grammar and vocabulary, Spanish and Basque seem to

express emotions with the same accuracy, not only in general terms, but they also showed similar recognition patterns. For example, Fear and Disgust were the worst recognized emotions in both Spanish and Basque. However, the fact that both languages were able to transmit emotions by means of vocal cues does not mean that such cues were identical. A matter of future work is to verify this by means of contrasting the utterance in function of the acoustic parameters.

## 5 Conclusions and Future Work

At the moment, the validated database is being used for training affective recognition applications applied to the cultural characteristics of the place where authors carried out their research. It is considered that training affective recognizers with subject validated databases will enhance the effectiveness of recognition applications. For example, the naturalness of the resources contained in the database is being analyzed taking voice parameters that have influence in the affective expression and recognition into account. Standard signal processing techniques have been used for extracting parameters involved in emotional speech. Several Machine Learning techniques have been applied to evaluate their utilities in the affective speech recognition [15]. The aim has been enhancing the results which are obtained with more traditional approaches

The aforementioned effectiveness must also take multimodal database features (images, linguistic parameters, vocabulary, etc.) into account. In the future, this database will be extended with combined recordings of audio and video.

Moreover, in future studies, social and contextual information will be added. All of this information will be described in an ontology with the aim of associating multimodal elements. Using this ontology in combination with software engineering, applications will assist in the development of affective systems [26], both for the scientific community and for industry.

Finally, another matter of future work is to contrast whether emotion transmission by vocal cues in Spanish and Basque are based on different or similar acoustic parameters.

**Acknowledgements.** The work carried out received financial support from the Department of Economy of the local government “Gipuzkoako Foru Aldundia” and from the University of the Basque Country (in the University-Industry projects area). The authors would like to express their gratitude to the people involved in the RekEmozio project in general and to the people that participated in the compilation and validation of the RekEmozio database in particular.

## References

1. Casacuberta, D.: La mente humana: Diez Enigmas y 100 preguntas (The human mind: Ten Enigmas and 100 questions). Océano (ed.), Barcelona, Spain (2001)
2. Garay, N., Abascal, J., Gardeazabal, L.: Mediación emocional en sistemas de Comunicación Aumentativa y Alternativa (Emotional mediation in Augmentative and Alternative Communication systems). Revista Iberoamericana de Inteligencia Artificial (Iberoamerican journal of Artificial intelligence) 16, 65–70 (2002)

3. Picard, R.W.: Affective Computing. MIT Press, Cambridge, MA (1997)
4. Tao, J., Tan, T.: Affective computing: A review. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 981–995. Springer, Heidelberg (2005)
5. Garay, N., Cearreta, I., López, J.M., Fajardo, I.: Assistive technology and affective mediation. Human technology. Special Issue on Human Technologies for Special Needs 2(1), 55–83 (2006)
6. Fragopanagos, N.F., Taylor, J.G.: Emotion recognition in human-computer interaction. Neural Networks 18, 389–405 (2005)
7. Athanaselis, T., Bakamidis, S., Dologlou, I., Cowie, R., Douglas-Cowie, E., Cox, C.: ASR for emotional speech: clarifying the issues and enhancing performance. Neural Networks 18, 437–444 (2005)
8. Cowie, R., Douglas-Cowie, E., Cox, C.: Beyond emotion archetypes: Databases for emotion modelling using neural networks. Neural Networks 18, 371–388 (2005)
9. Ekman, P., Friesen, W.: Pictures of facial affect. Consulting Psychologist Press, Palo Alto, CA (1976)
10. Mehrabian, A., Russell, J.A.: An approach to environmental psychology. MIT Press, Cambridge, MA (1974)
11. Tellegen, A.: Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Tuma, A.H., Maser, J.D. (eds.) Anxiety and the anxiety disorders, pp. 681–706. Lawrence Erlbaum, Hillsdale, NJ (1985)
12. Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., Gielen, S.: Acoustic correlates of emotion dimensions in view of speech synthesis. In Proc. Eurospeech 1, 87–90 (2001)
13. Navas, E., Hernáez, I., Castelruiz, A., Luengo, I.: Obtaining and Evaluating an Emotional Database for Prosody Modelling in Standard Basque. Lecture Notes on Artificial Intelligence, vol. 3206, pp. 393–400. Springer, Berlin (2004)
14. Iriondo, I., Guaus, R., Rodríguez, A., Lázaro, P., Montoya, N., Blanco, J.M., Bernadas, D., Oliver, J.M., Tena, D., Longhi, L.: Validation of an acoustical modelling of emotional expression in Spanish using speech synthesis techniques. In: SpeechEmotion'00, pp. 161–166 (2000)
15. Álvarez, A., Cearreta, I., López, J.M., Arruti, A., Lazkano, E., Sierra, B., Garay, N.: Feature Subset Selection based on Evolutionary Algorithms for automatic emotion recognition in spoken Spanish and Standard Basque languages. In: Sojka, P., Kopecek, I., Pala, K. (eds.) TSD 2006. LNCS (LNAI), vol. 4188, pp. 565–572. Springer, Heidelberg (2006)
16. López, J.M., Cearreta, I., Garay, N., de López Ipiña, K., Beristain, A.: Creación de una base de datos emocional bilingüe y multimodal. In: Redondo, M.A., Bravo, C., Ortega, M. (eds.) Proceedings of the 7th Spanish Human Computer Interaction Conference, Interaccion'06, Puertollano, pp. 55–66 (2006)
17. Pérez, M.A., Alameda, J.R., Cuetos Vega, F.: Frecuencia, longitud y vecindad ortográfica de las palabras de 3 a 16 letras del diccionario de la lengua española (RAE, 1992). Revista Española de Metodología Aplicada 8(2), 1–20 (2003)
18. Arrue, M., Fajardo, I., López, J.M., Vigo, M.: Interdependence between technical web accessibility and usability its influence on web quality models. Int. J. Web Engineering and Technology 3(3), 307–328 (2007)
19. López, J.M.: Development of a tool for the Design and Analysis of Experiments in the Web. In: Lorés, J., Navarro, R. (eds.) Proceedings of The 5th Spanish Human Computer Interaction Conference, Interacción'04, Lleida pp. 434–437 (2004)

20. Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T.: Vocal cues in emotion encoding and decoding. Motivation and Emotion 15, 123–148 (1991)
21. Banse, R., Scherer, K.R.: Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology 70(3), 614–636 (1996)
22. Johnstone, T., Scherer, K.R.: Vocal communication of emotion. In: Lewis, M., Haviland, J. (eds.) Handbook of Emotion, 2nd edn. pp. 220–235. Guilford Publications, New York (2000)
23. Abelin, A.: Cross-cultural multimodal interpretation of emotional expressions – an experimental study of spanish and swedish. In: Abelin, A. (ed.) SProSIG (2004)
24. Oudeyer, P.-.Y.: The production and recognition of emotions in speech: features and algorithms. International Journal of Human-Computer Studies 59(1-2), 157–183 (2003)
25. Tickle, A.: English and Japanese speaker's emotion vocalizations and recognition: a comparison highlighting vowel quality. In: ISCA Workshop on Speech and Emotion, Belfast (2000)
26. Obrenovic, Z., Garay, N., López, J.M., Fajardo, I., Cearreta, I.: An ontology for description of emotional cues. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 505–512. Springer, Heidelberg (2005)