# Analysis of Expressiveness of Portuguese Sign Language Speakers

Inês V. Rodrigues[1], Eduardo M. Pereira[1,2(✉)], and Luis F. Teixeira[1]

[1] Faculty of Engineering of the University of Porto,
Rua Dr. Roberto Frias, 378, 4200-465 Porto, Portugal
[2] INESC TEC, Porto, Portugal
ejmp@inescporto.pt

**Abstract.** Nowadays, there are several communication gaps that isolate deaf people in several social activities. This work studies the expressiveness of gestures in Portuguese Sign Language (PSL) speakers and their differences between deaf and hearing people. It is a first effort towards the ultimate goal of understanding emotional and behaviour patterns among such populations. In particular, our work designs solutions for the following problems: (i) differentiation between deaf and hearing people, (ii) identification of different conversational topics based on body expressiveness, (iii) identification of different levels of mastery of PSL speakers through feature analysis. With these aims, we build up a complete and novel dataset that reveals the duo-interaction between deaf and hearing people under several conversational topics. Results show high recognition and classification rates.

## 1 Introduction

The research on emotion, behaviour and expressiveness analysis was leveraged on 1962 by the important work on facial expression by Tomkins [1] and continued by Ekman [2] on 1975. The power of nonverbal behaviour in emotions became a central issue in most psychology textbooks as these started to be invaded by photos with prototypical expressions and simple emotions [3]. Nowadays, the panorama regarding the nonverbal emotional experiences has changed drastically. Every year, more than 50 books and papers are published featuring nonverbal channels of expressive communication. The channels considered are mainly facial expression, gestures, gaze, vocal quality, paralinguistic features, posture and body position, head nods, among others [3]. The focus of this work will be in body gestures which serve as main communicative function and contain substantial affective and cognitive information that help us emphasise certain parts of our speech and pass on expressive content.

Concerning the computer vision field, using automatic tools to extract body features allows a better understanding of the human behaviour so that it is possible to detect and interpret nonverbal temporal patterns. The fusion of several elements such as motion, appearance, shape, among others, is the key for solving the analysis of expressiveness that has driven the efforts of many researchers [4].


R. Paredes et al. (Eds.): IbPRIA 2015, LNCS 9117, pp. 708–717, 2015.
DOI: 10.1007/978-3-319-19390-8_79

Sign language speakers experience their languages very passionately. This may be explained by the fact that language plays a crucial role in the construction of a community and that it is a clear mark of belonging [5]. Emotion recognition from body language and its implications to the social adjustment of a sign language speaker are, therefore, very important issues. This study brings, in a novel way, this thematic to the field of computer vision, and presents important contributions: (i) creation of a novel video dataset that presents dialogues between deaf and hearing people (it will be release publicly for research purposes), (ii) proof of the differences in motion expressiveness between deaf and hearing people speaking PSL, (iii) the ability to distinguish different conversational moments that reveals behaviour differences based on context.

## 2 Related Work

Computer vision research on human behaviour analysis includes a broad range of studies to understand nonverbal sensitivity in different contexts and through different channels. Whole-body expressions provide information about the emotional state of the producer, but also signal his action intentions. The work in body expression was initiated in 1872 by Darwin who described the body expressions of many different emotions [6]. More recent studies showed that even in the absence of facial and vocal cues, it is possible to identify basic emotions signalled by static body postures [7], arm movement [8] and whole body movement [7–9]. Therefore, gestures serve an important communicative function in face-to-face communication since they often occur in conjunction with speech. According to Cassell [10], the fact that gestural errors are extremely rare demonstrates how essential their nature is for accurate communication.

For the particular case of the proposed work it is important to assimilate that gestures and the way they are performed work as an identity of a subject. The same action may be executed in several different ways depending on the executant. It is not rare that we are able to recognise a person by gait analysis, and it is also possible to infer emotional states by the way that person is moving [11]. This perspective enables us to classify walking as an expressive gesture [12]. Indeed, several everyday actions may constitute expressive gesture. For instance, Pollick et al. [13] investigated expressive content of actions such as knocking or drinking, and Heloir and Gibet [14] worked on the identification and representation of the variations induced by style for the synthesis of realistic and convincing expressive gesture sequences in sign language speakers. We bring this new topic to the field of expressiveness analysis in computer vision.

## 3 Methodology

As stated previously, emotion recognition from body language and its implications to the social adjustment of a sign language speaker are very important issues. Therefore, this study focused on evaluating the differences between deaf

and hearing people, in terms of expressive patterns through body motion analysis in several conversational topics. It is a preliminary work that intends to point out directions for future research that help to reduce the gaps that nowadays prevent deaf people from interacting easily with other people in society.

### 3.1 PSL Database

There was the need to create a database that could accomplish the requirements needed to accurately study differences of sign language speakers and measure their levels of nonverbal expressiveness according to the context. Indeed, the aim of the dataset was to enable the possibility of performing studies that analyse dialogue relationships between two individuals.

**Table 1.** Volunteer population for the creation of the database. All subjects are females between the ages of 27 and 39.

| Deaf people | Gender | Age | Hearing people | Gender | Age |
|---|---|---|---|---|---|
| D1 | Female | 38 | H1 | Female | 38 |
| D2 | Female | 35 | H2 | Female | 27 |
| D3 | Female | 36 | H3 | Female | 30 |
| D4 | Female | 39 | H4 | Female | 29 |
| D5 | Female | 31 | H5 | Female | 30 |
| D6 | Female | 39 | H6 | Female | 37 |
| – | – | – | H7 | Female | 37 |

We were advised by experts in the fields of social-psychology and sign-language. The conversational scenarios designed by the team of the Faculdade de Psicologia e Ciências da Educação da Universidade do Porto were: (i) conversation between two hearing people, (ii) conversation between two deaf people, (iii) conversation between a deaf and a hearing person. They also defined some requirements regarding the type of population and the recording scenario, for instance participants should have some kind of a priori knowledge between them, they should have the same gender, and the acquisition should take place in a venue that was familiar to all subjects. The contact with the volunteers was obtained by a partnership with the Agrupamento de Escolas Eugenio de Andrade, Escola EB2/3 de Paranhos. Table 1 displays the population that volunteered for the creation of this database.

Since our aim is the analysis of expressiveness, a set of conversational topics, that would awaken certain positive and negative emotions in the individuals, was defined. Those topics follow a staggered way so that the discussion would generate emotions of increasing intensity in the actors. The topics were chosen assuming that a dialogue would occur between a pair of subjects and that both would intervene actively. The conversation topics as well as the whole database

structure is detailed on Fig. 1. Each conversational pair of subjects is called a session. Since this work is oriented to the study of sign language speakers, the recording sessions in which two hearing people were having a conversation were not considered since they did not use sign language. We also discarded one session (between subjects D3 and D2) since they kept a standing position, which does not allow a fair comparison. Under this constraints, we base our experiences on 9 sessions.

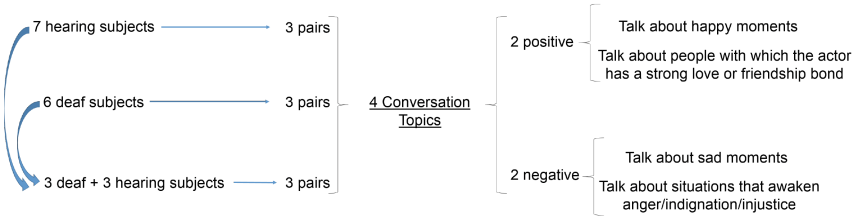


**Fig. 1.** Diagram of the database structure.

### 3.2 Dataset Preparation and Feature Construction

Each session is represented by a video and on each one a region of interest (ROI) was defined so that only the area bounding the subject was considered. Before building the feature vector for each video the following statistics were measured for each one: frames per second (fps), total number of frames and duration. Each one was subdivided into miniclips with the same number of frames (120) and the same fps (25) in order to be used as samples for the learning and classification tasks. Depending on each video's fps, some videos underwent a downsampling and others an up-sampling process. For each video, a region of interest (ROI) was defined so that only the area bounding the subject was considered.

We explored several trajectory and pixel based features to capture and represent body expressiveness regarding our aims. Some of those features were motion history image (MHI), motion gradients, several body part trackers, and other kinematic features. However, for the sake of simplicity and lack of space we just present here the results and conclusions obtained considering motiongrams feature [15]. Indeed, we performed some feature selection techniques that clearly highlight their discriminative power over the remaining features. Therefore, we use histogram representations for both vertical and horizontal motiongrams and concatenate them to be used as the final feature vector. Special care was taken in order to allow all feature vectors for each miniclip to have the same dimension. In this way, we considered the minimum values for width and height of all miniclips to reduce bin widths, obtaining a 114 and 162 dimensional bin size for both vertical and horizontal motiongrams, respectively.

### 3.3 Distinguishing Deaf from Hearing People

For this problem the two classes were known a priori for each miniclip, namely if the performer was a deaf or a hearing person. For classification purposes, the miniclips were grouped by topic so that it was possible to compare the classification performance for the different groups. We considered two classification methods: $k$-Nearest Neighbours ($k$-NN), which is simple and widely used as a first approach for classification problems, and Support Vector Machine (SVM), which is a more complex algorithm but usually more accurate. Both algorithms were used under a cross-validation mechanism and different number of folds (dependent on the number of samples per grouping) were used in order to avoid over-fitting. For the SVM we performed a grid search to automatically infer the optimal parameters. The statistical measures used to evaluate the performance of these two classifiers were the Confusion Matrix (CM), Correct Rate (CR), Recall (R) and Precision (P).

### 3.4 Distinguishing Different Conversation Topics

Regarding the differentiation of the conversational topics (two with positive connotation and two with negative) spotting the differences in terms of expressiveness among the four topics was the primary reason why these moments were included on the database. In this case, the miniclips were grouped by subject.

The same classifiers, $k$-NN and SVM, were used to approach this problem. However, we use a multi-class SVM for ordinal data, since we want to inspect if the topics could follow a natural order of expressiveness, and analyse the intra-class and inter-class boundary decision between the topics that belong to the same positive or negative connotation. We used the approach generalised in [16]. The same metrics were used for evaluation performance.

### 3.5 Identifying Levels of Mastery in PSL

For the purpose of distinguishing different levels of expertise in PSL, an agglomerative hierarchical clustering method was used. In order to identify groups of similar feature values, clustering procedures use distance measures to group data points in a way that provides minimal inner-cluster distances and maximal inter-cluster distances [17]. The possibility of our miniclips containing information regarding this question was not known a priori, this means that we were not aware if it would be possible to obtain a reliable answer. This problem is very abstract but also a valuable addition to the overall framework of this study.

One of the drawbacks of the agglomerative approach is the requirement for the number of clusters to be specified before the algorithm is applied. In order to overcome this issue, using the questionnaires that all subjects filled, it was decided to define 3 and 4 levels of expertise in PSL based on the score of each subject's answers to the following questions: number of years familiarised with sign language and the current profession. The combined score of the two answers was also used. Table 2 shows the organisation of the subjects in levels regarding

their answers to the two mentioned questions. The weighted combination of the scores of each answer (0.5 for each) originated 4 different levels.

**Table 2.** Division of our population regarding the number of years in contact with the PSL and current job for classification purposes.

| Level | Year range | Subjects | Level | Current profession | Subjects |
|---|---|---|---|---|---|
| 1 | 7–15 | D1 / H2 / H7 | 1 | Non Related to PSL | D4 |
| 2 | 16–25 | H6 / D5 / D6 | 2 | Speech Therapist | H7 |
| 3 | 26–35 | D4 | 3 | PSL Interpretation | H2 |
| – | – | – | 4 | PSL Teaching | D1 / D5 / D6 / H6 |

## 4 Results and Discussion

### 4.1 Distinguishing Deaf from Hearing People

Distinguishing deaf from hearing people when using PSL is the main enquiry to be answered by this work. Table 3 shows the results of the two classifiers used for the different data groupings.

**Table 3.** $k$-NN and SVM classification results obtained for the task of distinguishing between deaf and hearing people.

| | Correct rate | | Recall | | Precision | |
|---|---|---|---|---|---|---|
| | $k$-NN | SVM | $k$-NN | SVM | $k$-NN | SVM |
| Topic 1 | 0.98 | 0.98 | 0.98 | 0.98 | **0.99** | 0.97 |
| Topic 2 | 0.95 | **0.98** | 0.97 | **0.98** | 0.94 | **0.98** |
| Topic 3 | 0.97 | **0.98** | 0.97 | **0.99** | 0.98 | **0.99** |
| Topic 4 | 0.97 | **0.98** | 0.98 | 0.98 | 0.97 | **0.98** |
| All topics | 0.96 | **0.97** | **0.97** | 0.96 | 0.96 | 0.96 |

The results obtained for both classifiers are in concordance with each, although the performance is slightly improved for the SVM. For the SVM the best performance is observed for Topic 3. Inspecting Table 4, we observe a balance in the misclassification rates of the classes which leads to conclude that both classifiers perform well, and that our feature vector is highly discriminative for all classes. Accumulated CM are more a less equivalent for all the miniclip groupings, being the highest misclassification rates observed for the grouping of all topics, where more samples may confuse the classifier.

It is reasonable to say that the information contained in the videos of our database is rich and the feature approach made to extract that information was

**Table 4.** Accumulated confusion matrix (CM) for both $k$-NN and SVM methods.

| | $k$-NN | CM | SVM | CM |
|---|---|---|---|---|
| Topic 1 | 0.97 | 0.03 | 0.97 | 0.03 |
| | 0.05 | 0.95 | 0.01 | 0.99 |
| Topic 2 | 0.95 | 0.05 | 0.98 | 0.02 |
| | 0.04 | 0.96 | 0.00 | 1.00 |
| Topic 3 | 0.98 | 0.02 | 0.99 | 0.01 |
| | 0.04 | 0.96 | 0.01 | 0.99 |
| Topic 4 | 0.98 | 0.02 | 1.00 | 0.00 |
| | 0.03 | 0.97 | 0.03 | 0.97 |
| All topics | 0.97 | 0.03 | 0.98 | 0.02 |
| | 0.05 | 0.95 | 0.04 | 0.96 |

appropriate. The consulted PSL experts stated that in order to distinguish if a subject using LGP is deaf or hearing we should focus mainly on evaluating facial features. With this result we prove that the body features are also very descriptive.

### 4.2 Distinguishing Different Conversation Topics

When the database was created, the definition of four different moments had the goal of making it more robust and complete. Only PSL speakers are able to evaluate the contents of the videos and verify if the four distinct moments were in fact present. Therefore, this supervised evaluation was done for some of the videos confirming that the subjects were demonstrating different levels of expressiveness and emotion accordingly to the current discussion topic.

The remaining videos were also analysed without expert supervision in order to extract some conclusions based on the queues given previously by the experts. It was possible to deduce that, regarding the presence of four different conversational moments, the framework being developed is accurate and valuable. Table 5 shows the results of the two classifiers used for the different data groupings.

### 4.3 Identifying Levels of Mastery in PSL

In this task we used a non-supervised agglomerative hierarchical tree to build a hierarchy of clusters. Considering the process explained in Sect. 3.5, we get different configurations for the number of clusters to be used: (i) 3 in the case that we wanted to group our subjects by the number of years familiarised with PSL, (ii) 4 for the current profession, (iii) 4 for the combined score of the answer of the two questions. Tables 6, 7 and 8 show the evaluation statistics of the clusterings performed on each configuration, respectively, where *cid* indicates the id of the cluster, *Size* the number of samples that belong to each cluster, *ISim* and *ISdev* represent both the average and standard deviation in terms of

**Table 5.** *k*-NN and SVM classification results obtained for the task of distinguishing the different conversational topics.

| | Correct rate | | Recall | | Precision | |
|---|---|---|---|---|---|---|
| | k-NN | SVM | k-NN | SVM | k-NN | SVM |
| Subject D1 | 0.91 | 0.91 | 0.85 | **0.92** | 0.90 | 0.90 |
| Subject D4 | 0.96 | **0.98** | 0.94 | **0.98** | 0.91 | **0.97** |
| Subject D5 | **0.93** | 0.91 | 0.86 | **0.90** | **0.93** | 0.89 |
| Subject D6 | **0.93** | 0.90 | **0.95** | 0.89 | **0.96** | 0.89 |
| Subject H2 | **0.91** | 0.86 | **0.93** | 0.86 | **0.98** | 0.85 |
| Subject H6 | 0.89 | **0.98** | 0.83 | **0.98** | 0.89 | **0.97** |
| Subject H7 | 0.96 | **0.97** | 0.94 | **0.98** | **0.99** | 0.98 |
| All subjects | **0.92** | 0.78 | **0.91** | 0.82 | **0.93** | 0.77 |

similarity between each cluster, whereas *ESim* and *ESdev* represent the same statistics but for similarity of the objects of each cluster and the rest of the objects.

**Table 6.** Clustering statistics: the class considered is the number of years in contact with PSL.

| cid | Size | ISim | ISdev | ESim | ESdev | Entpy | Purty | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 70 | 0.669 | 0.075 | 0.332 | 0.079 | 0.625 | 0.557 | 0.44 | 0.56 | 0.00 |
| 2 | 167 | 0.606 | 0.105 | 0.477 | 0.117 | 0.596 | 0.707 | 0.71 | 0.28 | 0.01 |
| 3 | 1029 | 0.597 | 0.112 | 0.437 | 0.12 | 0.75 | 0.646 | 0.28 | 0.65 | 0.07 |

According to the opinion of the PSL experts, the number of years alone might not be a clear indicator of how experienced a subject is in terms of PSL, since several factors may influence it. In fact, the overall entropy and purity of this clustering are 0.723 and 0.649 respectively. These values reveal that the clustering was performed poorly. Purity is quite low which indicates us that the samples in each cluster are not as homogeneous as desired.

**Table 7.** Clustering statistics: the class considered is the current profession of the subject.

| cid | Size | ISim | ISdev | ESim | ESdev | Entpy | Purty | 4 | 3 | 1 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 70 | 0.669 | 0.075 | 0.332 | 0.079 | 0.584 | 0.6 | 0.60 | 0.36 | 0.00 | 0.04 |
| 2 | 167 | 0.606 | 0.105 | 0.477 | 0.117 | 0.529 | 0.695 | 0.69 | 0.28 | 0.01 | 0.02 |
| 3 | 144 | 0.709 | 0.078 | 0.559 | 0.087 | 0.669 | 0.674 | 0.67 | 0.12 | 0.18 | 0.03 |
| 4 | 885 | 0.602 | 0.111 | 0.481 | 0.127 | 0.536 | 0.739 | 0.74 | 0.01 | 0.05 | 0.20 |

When performing the clustering in relation to the levels in terms of professional occupation, the overall entropy and purity of the solution were the best of the three analysis performed (0.553 and 0.718). From the questions featured on the questionnaires, this one regarding the current professional occupation of the subjects was the one considered by PSL experts to possibly be more discriminative when it comes to the expertise on this language. This is confirmed by an improvement on entropy and purity values.

**Table 8.** Clustering statistics: the class are the number of years in contact with PSL combined with the current profession.

| cid | Size | ISim | ISdev | ESim | ESdev | Entpy | Purty | 3 | 4 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 70 | 0.669 | 0.075 | 0.332 | 0.079 | 0.695 | 0.557 | 0.04 | 0.56 | 0.36 | 0.04 |
| 2 | 167 | 0.606 | 0.105 | 0.477 | 0.117 | 0.831 | 0.413 | 0.41 | 0.28 | 0.29 | 0.02 |
| 3 | 144 | 0.709 | 0.078 | 0.559 | 0.087 | 0.845 | 0.424 | 0.25 | 0.42 | 0.30 | 0.03 |
| 4 | 885 | 0.602 | 0.111 | 0.481 | 0.127 | 0.66 | 0.682 | 0.06 | 0.68 | 0.06 | 0.20 |

The combination of previous classes into a weighted distribution was done with the intention of obtaining classes that could concatenate more information about the expertise of each subject. The less promising results of entropy and purity tell us this fusion impair the performance of the clustering solution. To overcome this issue a different type of combination of the information could help.

## 5 Conclusions

This study was focused on automated visual analysis of expressiveness of PSL speakers using computer vision techniques. We achieve important breakthroughs under this topic: (i) we provide a novel and rich dataset for the study of expressiveness and emotion states on a duo-interaction between deaf and hearing people, (ii) we achieve a high discriminative feature vector capable of distinguishing deaf from hearing people and also different conversational topics, (iii) we point out directions about the stratification of levels of expertise in PSL, and their recognition through the correlation of video with social data. The future work intends to go in the direction of finding emotional and behaviour patterns through face and body expressions analysis, of deaf and hearing people. New techniques and approaches need to be reviewed and tested since this is a very ambitious goal.

**Acknowledgment.** The authors would like to thank to the PSL experts, Ana and Paula, who helped to find the volunteer population, to the socio-psychologist team from the Faculdade de Psicologia e Ciências da Educação da Universidade do Porto who helped to define the sociological constraints of the database, to the Agrupamento de Escolas Eugenio de Andrade, Escola EB2/3 de Paranhos for providing the venue for acquisition of the videos of the database, and finally to Stephano Piana for his help with the EyesWeb platform.

## References


1. Tomkins, S.: Affect Imagery Consciousness: Volume:II: The Negative Affects. Springer Series. Springer Publishing Company, New York (1963)
2. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
3. Harrigan, J., Rosenthal, R., Scherer, K.: New Handbook of Methods in Nonverbal Behavior Research. Series in affective science. OUP, Oxford (2008)
4. Metaxas, D., Zhang, S.: A review of motion analysis methods for human nonverbal communication computing. Image Vis. Comput. **31**(6–7), 421–433 (2013). Machine learning in motion analysis: New advances
5. Nadal, J.M., Monreal, P., Perera, S.: Emotion and linguistic diversity. Procedia Soc. Behav. Sci. **82**, 614–620 (2013). World Conference on Psychology and Sociology 2012
6. Darwin, C.: The Expression of the Emotions in Man and Animals. John Murray, London (1872)
7. Nakajima, C., Pontil, M., Heisele, B., Poggio, T.: Full-body person recognition system. Pattern Recogn. **36**(9), 1997–2006 (2003)
8. Piana, S., Staglianó, A., Odone, A.C.A.: A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In: IDGEI International Workshop (2013)
9. Atkinson, A., Dittrich, W., Gemmell, A., Young, A.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception **33**, 717–746 (2004)
10. Cassell, J.: A framework for gesture generation and interpretation. In: Cipolla, R., Pentland, A. (eds.) Computer Vision in Human-Machine Interaction, pp. 191–215. Cambridge University Press, Cambridge (2000)
11. Kobayashi, Y.: The emotion sign: human motion analysis classifying specific emotion. JCP **3**(9), 20–28 (2008)
12. Hwang, B.-W., Kim, S.-M., Lee, S.-W.: 2D and 3D full-body gesture database for analyzing daily human gestures. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 611–620. Springer, Heidelberg (2005)
13. Pollick, F., Paterson, H., Bruderlin, A., Sanford, A.: Perceiving affect from arm movement. Cognition **82**(2), B51–61 (2001)
14. Heloir, A., Gibet, S.: A qualitative and quantitative characterisation of style in sign language gestures. In: Sales Dias, M., Gibet, S., Wanderley, M.M., Bastos, R. (eds.) GW 2007. LNCS (LNAI), vol. 5085, pp. 122–133. Springer, Heidelberg (2009)
15. Jensenius, A.R.: Using motiongrams in the study of musical gestures. In: Proceedings of the International Computer Music Conference, pp. 499–502. Tulane University, New Orleans (2006)
16. Pinto da Costa, J., Sousa, R., Cardoso, J.: An all-at-once unimodal svm approach for ordinal classification. In: Ninth International Conference on Machine Learning and Applications (ICMLA 2010), pp. 59–64, December 2010
17. Zhao, Y., Karypis, G.: Evaluation of hierarchical clustering algorithms for document datasets. In: Proceedings of the Eleventh International Conference on Information and Knowledge Management, CIKM 2002, pp. 515–524. ACM, New York (2002)