VirtualSign Game Evaluation

Paula Escudeiro, Nuno Escudeiro, Marcelo Norberto(✉),
and Jorge Lopes

Informatic Engeneering Department, Instituto Superior de Engenharia do Porto,
Porto, Portugal
{pmo,nfe,matno,jolbl}@isep.ipp.pt

Abstract. This paper presents the evaluation methods and techniques applied to the serious game developed within the VirtualSign project as well as the results achieved through those methods. VirtualSign is a Portuguese sign language bi-directional translator with three main components. The components are a gesture to text translator, a text to gesture translator and a serious game. The serious game aims to make the process of learning sign language easier and enjoyable using the VirtualSign bi-directional translator. The method used to evaluate the VirtualSign was the Quantitative Evaluation Framework (QEF). The translator undertakes a rigid validation process by both sign language experts and the deaf community. The evaluation process was also supported by questionnaires. The precision of the translator has a direct impact on the game performance. To evaluate the precision of the automatic translator we have used QEF and 10-fold cross validation to estimate the accuracy of the translator. The experimental results show a precision above 90%. In the identification of the Portuguese Sign Language terms As for the game the quality rate is 88% based on the QEF evaluation.

Keywords: VirtualSign · Serious game · Results · Sign language · Translator

1 Introduction

The interest in Portuguese Sign Language (PSL) has shown a remarkable growth over time, not only by the deaf community - which now accounts for nearly 100000 persons in Portugal - but also for the whole community involved, such as relatives, educators, teachers, and many more.

The Virtualsign project aims to improve the social inclusion of this community. In order to do so a bi-directional translator of PSL was created. It allows the translation from gestures to text using a pair of sensor gloves and the Microsoft Kinect which provides the necessary data for the sign recognition. Also, it allows the translation from text to gesture using an animated 3D avatar that performs the gestures received by text. After the creation of the translator an additional component was added to the project, a serious game.

Serious games provide a remarkable opportunity to overcome the lack of educational digital content available for the hearing impaired community. A well-designed game entices players into the “reality” of the game world and keeps them there until the

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2017
C. Vaz de Carvalho et al. (Eds.): SGAMES 2016, LNICST 176, pp. 117–124, 2017.
DOI: 10.1007/978-3-319-51055-2_15

goals of the game have been met [1]. Therefor there is a need to validate the performance and design of any serious game.

The method used to validate the Virtual sign game was the QEF [2], however to improve the evaluation process additional validations were performed by both sign language specialist and the deaf community through questionnaires. The Virtualsign game depends on the translator for both recognition of the gestures and the performing of the gestures through an avatar inside the game [3]. In the game the player will control a 3D avatar that will perform gestures corresponding to sign language and the user himself has specific areas where he will be prompt to perform a gesture. This dependency means that the game is affected directly by the bi-directional translator performance. The bi-directional translator uses the QEF for evaluation just as the game but has an additional technique named 10-fold cross-validation that estimates the accuracy and precision of the classifiers. The results of both the translator and the game account for the overall quality of the game.

2 Related Work

Despite the lack of content related to this area there are a few projects under development but there is yet very little information about them.

The most similar project to the one mentioned in this paper is the CopyCat. The project CopyCat [4] is an interactive adventure and educational game with ASL recognition. Colorful gloves equipped with accelerometers are used in order to simplify the segmentation of the hands and allow the estimation of motion acceleration, direction and the rotation of the hands. The data is classified using Hidden Markov Model (HMM) [5], yielding an accuracy of 83%.

Another application related to this work is ProDeaf [6], although it is not a serious game it is very similar to one of the main components used on the VirtualSign game, which is the text to gesture translation.

ProDeaf is an application that does the translation of Portuguese text or voice to Brazilian Sign language named Libras. The objective of the ProDeaf is to make the communication between mute and deaf people easier by making digital content accessible in Libras. The translation is done using a 3D avatar that performs the gestures. This software is already used by over 130 000 users.

There is also the application Handtalk [7] which is very similar to the prodeaf, it translates Portuguese text or voice to Libras using a 3D avatar. It's available for mobile and there is also a plugin for websites.

As for the sign recognition there are several models.

Celal Savur et al. [8] suggests a real-time American Sign Language recognition system using the surface Electromyography (EMG). Surface EMG data is acquired from subject's right forearm domain information. The experiment result of offline system is reaching a accuracy rate of 91.1% and real-time system has an accuracy rate of 82.3% using Support Vector Machine.

Nicolas Pugeault et al. [9] suggest a system for recognition of the hand configuration in the context of ASL, using the Microsoft Kinect to collect information about appearance and depth, and the OpenNI + NITE framework to detect and track the hand. The collected data is classified by applying a random forests algorithm, yielding an average accuracy rate of 49.5%.

3 Evaluation Methodology

3.1 Game Quantitative Evaluation Framework

QEF is a generic quality evaluation framework. This framework may also be applied in other settings. The QEF evaluates the educational software quality (ISO 9126 is the standard of reference) Scalet et al. [10] in a three dimensional space. A dimension aggregates a set of factors. A factor is a component that represents the system performance from a particular point of view. The dimensions of the Virtualsign Game Cartesian quality space are: Pedagogical (P); Ergonomic (E) and Technical (T), represented in Fig. 1.

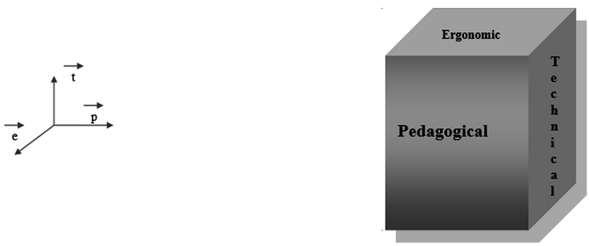

Fig. 1. Cratesian quality space of the game

The quality q, of a given system is defined in our tri-dimensional Cartesian quality space, Q, and measured, in percentage, relatively to a hypothetically ideal system, I.

The Pedagogical dimension reflects the characteristics of the educational software related to its learning aspects. It aggregates two factors: learning and assessment.

The Ergonomic dimension aggregates two factors: usability and gameplay. Those factors measure the system's ability of presenting its content with minimum effort.

The Technical dimension for this case only has one factor which is the Support.

For each system being developed we will have to identify the importance of each factor to the dimension, pn in Eq. 1. The dimension coordinate is them computed as the weighted mean of these factors:

$$\mathrm{Dim_i} = \sum_{n} (p_n \times factor_n),\ \sum_{n} (p_n) = 1 \mathrm{e}\, p_n \in [0, 1] \tag{1}$$

Where n is the number of relevant factors for the dimension at hand.

The ideal system has a set of requirements that indicates what the system must do. Those requirements have an associated weight according to a scale from 0 to 10 where 0 is irrelevant and 10 fundamental. The weight of a given criterion is the relevance of the criterion to the factor, which have been assigned previously. These weights define the ideal system.

$$\text{Factor}_{n=} \frac{1}{\sum_m pr_m} \times \sum_m (pr_m \times pc_m) \tag{2}$$

In the Eq. 2 pr_m is the weight of the criterion m for the factor under evaluation. The pc_m represents the fulfillment percentage of the requirement m.

The global deviation of the real system – represented by the three dimensional coordinates in the quality space – with regards to the ideal system – represented by the coordinates (1,1,1) – is computed by the Euclidean distance between these two points in quality space.

The real system quality is then computed as seen in Eq. 3.

$$Q = 1 - \frac{D}{\sqrt{n}}, \ Q \in [0, 1] \tag{3}$$

QEF has previously been applied to control the quality of several products throughout their lifecycle with very good results.

3.2 VirtualSign Translator Evaluation

The Virtualsign bi-directional translator was also evaluated using the QEF. However, the set of dimensions used for the translator were Functionality, Adaptability and Efficiency. The 10-Fold Cross-Validation was used in order to estimate the accuracy of the classifiers responsible for the analysis and classification of the gestures. The cross-validation is a model validation technique used to assess how the results of a statistical analysis vary according to an independent data set. The model is usually given a dataset of known data to be used in the training of the classifier (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (testing dataset). This validation allows the prediction of how accurate the classifiers are.

Besides the QEF and the cross-validation, several questionnaires were delivered to deaf users who tried the text to gesture component of the Virtualsign translator. Each user translated 275 words on the translator and replied whether the translation performed by the avatar was correct, wrong or they could leave a comment. The total number of deaf users who replied the questionnaires were four, two male and two female, and their ages are ranged from 39 to 64 years old.

4 Results

4.1 QEF Requirments and Results of the Game

The VirtualSign game main goal is to aid those who are willing to learn sign language. In order to do so its content had to be validated and the game design had to be the adequate for a serious game. The QEF was used for those validations along with other tools and experts. A set of requirements were chosen aiming to evaluate the main aspects that should be part of any serious game.

In Table 1 there is the list of requirements for the pedagogical dimension of the game. The learning factor has a weight of 0.75 and the assessment has a weight of 0.25. All the requirements are essential to the game and therefore all have the same weight of 10.

Table 1. Requirement list for the pedagogical dimension of the QEF.

Factors	Requirements
Learning	PL01 - The story line promotes interactive learning curve
	PL02 - The game promotes an incremental educational guide
	PL03 - The game elements represent clearly the educational objects in the real world
	PL04 - The game evaluates and awards the player that capture more educational content information
	PL05 - The checkpoints use clear language for the target group
	PL06 - In the end of the checkpoints the player is awarded by its performance
Assessment	PA01 - The game promotes self-assessment
	PA02 - The game provides a screen to check out the top player results

The ergonomic dimension has the most requirements as it contains the usability factor, which weight is 0.55 and the gameplay factor with a weight of 0.45. Those two factors are the core of any game thus having such a large amount of requirements associated with each. Those requirements are listed in Table 2.

Finally, the technical dimension of the QEF for the game is listed in Table 3, it contains only the support factor with a weight of 1.

The pedagogical dimension has a completion rate of 100% whilst the ergonomic is at 84.9% and the Technical at 70%.

In the ergonomic dimension the usability factor is at 78.1% as the game is missing the help button and difficulty adjustment. As for gameplay is at 93.2% lacking some audio feedback the pace could be improved. Finally the technical dimension is missing a verification upon a game crash to save the game state and the game is not easily updated thus the 70% only.

Based on the set of requirements for the game it has a quality percentage of 88% according to the QEF. The game performance also depends on the translator therefore the results of the translator have been also evaluated and described in this paper.

Table 2. Requirement list for the ergonomic dimension of the QEF.

Factors	Requirements
Usability	EU01 - The game is easy and does not require a large learning curve
	EU02 - The user should be able to retry a level to try to achieve a better result
	EU03 - A help button is provided
	EU04 - The game difficulty is increased by level
	EU05 - The common element's colors used in the game were consistent throughout it
	EU06 - Written/spoken content is free of grammatical and syntactical errors.
	EU07 - Feedback for users action is quick and effective
	EU08 - Main game menus and in game menus are easy to use
	EU09 - Options in game menu correctly do what their supposed to
	EU10 - A user can leave and restart the adventure (level) anytime during the gameplay
	EU11 - The controls of the game are easy to use
	EU12 - The graphics were recognizable to the players and it was clear what items were by looking at them.
	EU13 - All user input actions with active objects in game set result in audio feedback, visual feedback or both.
	EU14 - The color palate used is clear and lean avoiding health problem for people with epilepsy
	EU15 - The player can set an initial difficulty for the game
	EU16 - The game should provide a cooperative mode for multiplayer
Gameplay	EG01 - The game story is well represented in the game sets
	EG02 - The game is original
	EG03 - The effects of the player's actions can clearly be seen in the game environment
	EG04 - Audio usage enhances gameplay
	EG05 - The game pace is quick and pleasing
	EG06 - The game is challenging and defies the user to beat his own best results
	EG07 - Game content enhances the game play
	EG08 - Application allows the recognition of patterns
	EG09 - The characters design are representative of the real elements that they represent
	EG10 - Game navigation and actions capture is quick and fluid
	EG11 - The game story creates an educational immersive context for the players

Table 3. Requirement list for the techincal dimension of the QEF.

Factors	Requirements
Support	TS01 - The game is easy updatable
	TS02 - Usage statistics and user history is saved
	TS03 - There is a unique entry point to the game
	TS04 - In case of crash the game should be able to be set to the latest valid state without data loss
	TS05 - The connection with the VS translator is stable and fast

4.2 Results of the Translator

Based on the QEF the Virtualsign translator has a quality of 89%. The dimension with the lowest quality percentage is the adaptability with 66.6%. As for the efficiency it's at 100% and the functionality is at 92% as the misspellings and syntax correction system is not yet fully functional.

According to the questionnaires about the text to gesture translation, an average of 74% of the words were correct, 6% could use some improvements and 20% were incorrect, however, based on the comment section of the questionnaires it's possible to conclude that some of the incorrect words are synonyms or differ of the dialect of the person who answered.

As for the performance of the gesture to text component six classification algorithms were compared using 10-fold cross validation, namely Random Trees (RT) [11], Boost Cascade (BC) [12], Neural Networks (NN) [13], K-Nearest Neighbors (KNN) [14], Naive Bayes (NB) [15] and Support Vector Machine (SVM) [16]. To evaluate their performance a dataset composed of 40 samples was used for each hand configuration (1680 samples in total). In Table 4 the results of the evaluation are presented.

Although RT shows a slightly higher precision the SVM classifier became the default of the VirtualSign yielding a precision of 98,6 for the left hand and 98,1 for the right as in this test it obtained an accuracy of 100%.

Table 4. Classification results of the 1680 samples, obtained with the sensor gloves.

%	RT	BC	NN	KNN	NB	SVM
Precision Left	98,6	82,0	98,1	98,8	97,5	98,6
Accuracy Left	85,5	95,4	78,1	97,3	97,1	100,0
Precision Right	98,8	86,1	97,2	98,0	98,0	98,1
Accuracy Right	87,3	96,6	80,4	98,2	96,8	100,0

5 Conclusion

According to the QEF the game shows a quality rate of 88%. The game is complete but a multiplayer cooperation system is under development for it and also the fact its performance is influenced directly by the translator makes it hard for the game to reach the 100% quality rate that we are aiming for. However even though the translator precision is not 100% it's still higher than any current alternatives. For instance the CopyCat is currently the only similar game and its translator shows a lower accuracy rate than the Virtualsign. The Virtualsign translator accuracy overall stands above 90% which is a very impressive result compared to the existing models. Both the translator and the game are being continuously improved to ensure the best user experience. So far the feedback from those who tested it has been very positive and many of the improvements came from it. As for future work there is a cooperation module under development to be added to the game. There is also in progress a parameterization application that will fill the gesture database using very limited amounts of data which will improve the game performance. We are also looking for alternatives to the sensor gloves in order to improve the usability of the game.

References

1. Salen, K., Zimmerman, E.: Rules of Play: Game Design Fundamentals. MIT Press, Cambridge (2004)
2. Escudeiro, P., Bidarra, J.: Quantitative evaluation framework (QEF). Conselho Editorial/ Consejo Editorial, p. 16 (2008)
3. Escudeiro, P., Escudeiro, N., Norberto, M., Lopes, J.: VirtualSign in serious games. In: de Carvalho, C.V., Escudeiro, P., Coelho, A. (eds.) Serious Games, Interaction, and Simulation, pp. 42–49. Springer, Heidelberg (2016)
4. Zafrulla, Z., Brashear, H., Yin, P., Presti, P., Starner, T., Hamilton, H.: American sign language phrase verification in an educational game for deaf children. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 3846–3849. IEEE (2010)
5. Blunsom, P.: Hidden markov models. Lecture notes, pp. 18–19, 15 August 2004
6. Prodeaf. http://www.prodeaf.net/en-us/Solucoes
7. HandTalk. https://www.handtalk.me/sobre
8. Savur, C., Sahin, F.: Real-time american sign language recognition system using surface EMG signal. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 497–502. IEEE, December 2015
9. Pugeault, N., Bowden, R.: Spelling it out: real-time ASL fingerspelling recognition. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1114–1119. IEEE, November 2011
10. Scalet, D., Yokohama, A., Koscianski, A., Rêgo, C.M., Asanome, C., Romero, D., Vostoupal, T.M.: ISO/IEC 9126 and 14598 integration aspects: a Brazilian viewpoint. In: The Second World Congress on Software Quality (2000)
11. Le Gall, J.F.: Random trees and applications. Probab. Surv. **2**(245), 10–1214 (2005)
12. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, pp. 1–511 (2001)
13. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall, Upper Saddle River (2004)
14. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theor. **13**(1), 21–27 (1967)
15. Rish, I.: An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, pp. 41–46. IBM, New York (2001)
16. Steinwart, I., Christmann, A.: Support vector machines. In: Izenman, A.J. (ed.) Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer Texts in Statistics. Springer, New York (2008)