Recognizing an Individual, Their Topic of Conversation, and Cultural Background from 3D Body Movement

Gheida J. Shahrour, Martin J. Russell

Abstract—The 3D body movement signals captured during human-human conversation include clues not only to the content of people's communication but also to their culture and personality. This paper is concerned with automatic extraction of this information from body movement signals. For the purpose of this research, we collected a novel corpus from 27 subjects, arranged them into groups according to their culture. We arranged each group into pairs and each pair communicated with each other about different topics.

A state-of-art recognition system is applied to the problems of person, culture, and topic recognition. We borrowed modeling, classification, and normalization techniques from speech recognition. We used Gaussian Mixture Modeling (GMM) as the main technique for building our three systems, obtaining 77.78%, 55.47%, and 39.06% from the person, culture, and topic recognition systems respectively. In addition, we combined the above GMM systems with Support Vector Machines (SVM) to obtain 85.42%, 62.50%, and 40.63% accuracy for person, culture, and topic recognition respectively.

Although direct comparison among these three recognition systems is difficult, it seems that our person recognition system performs best for both GMM and GMM-SVM, suggesting that inter-subject differences (i.e. subject's personality traits) are a major source of variation. When removing these traits from culture and topic recognition systems using the Nuisance Attribute Projection (NAP) and the Intersession Variability Compensation (ISVC) techniques, we obtained 73.44% and 46.09% accuracy from culture and topic recognition systems respectively.

Keywords—Person Recognition, Topic Recognition, Culture Recognition, 3D Body Movement Signals, Variability Compensation.

I. INTRODUCTION

SINCE Bolt's direct manipulation interface 'Put That There' in the early 1980s [1], researchers started to analyze human body movements. These movements contain a rich source of information (e.g. topic cues, culture specific cues, and the human personality traits) which can be a blessing to one task and a curse to another. In this paper, we are interested in this source of information for topic, culture, and person recognition. Specifically, we are interested in removing the information that degrades topic and culture recognition.

To conduct our research, we need a novel corpus that

G. J. Shahrour is a doctoral research student at the School of Electronic, Electrical and Systems Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK (phone: +44 (0) 121 414 2825; e-mail: gjs607@bham.ac.uk).

M. J. Russell is with the School of Electronic, Electrical and Systems Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK (phone: +44 (0) 121 414 3093; e-mail: m.j.russell@bham.ac.uk).

contains body movements signals captured from at least two cultural groups of individuals. The signals must be captured while a pair of individuals conversed with each other about various topics. The absence of such dataset led us to collect our own corpus.

We applied statistical techniques to build topic, culture and person recognition systems using training datasets from the corpus. In addition, we utilized techniques to remove the variations that degrade the accuracy of topic and culture recognition systems. To our knowledge, there is no existing research about building topic or culture recognition systems based solely on human body movement signals, and there is no research in the literature about suppressing irrelevant variations in topic and culture recognition systems.

II. CULTURE AND TOPIC RECOGNITION

Human body movements during conversation are a combination of co-speech gestures and other actions. The co-speech gestures (e.g. emblematic, iconic, deictic, and symbolic) [1] are linked, both in form and meaning, to the words they usually accompany [2], [3]. For example, we say 'tall', for example, while gesturing the 'tallness' of a tall person. Consequently, these gestures can contain cues that are useful for topic recognition. On the other hand, other actions that accompanied conversation (e.g. combing hair, tapping on the table, or flicking fingers) are not related, in general, to the content of the conversation. Consequently, these actions can degrade the accuracy of a topic recognition system. An interesting challenge is to analyse this combination of movements to see how well a topic recognition system performs under the condition where no speech is available.

Body movement signals contain a wealth of information over and above its topic cues or communicative intent, including clues (e.g. movement patterns or their frequency) to the ethnic background of the individual who is performing them [4], [5]. For example, instead of turning their head side to side as a British person would do to say 'no', Arabs moved their head up. In addition, Italians gesture more frequently during conversation than other Europeans [5]. Thus being products of people's cultural training [6], these movements can contain culture specific cues useful to culture recognition.

Most recognition systems are based on key features that relevant to the target task. For example, traditional topic recognition systems are based on language - be it written or spoken - where in some cases useful key words are specified beforehand and then queried [7]. Whereas, traditional culture

recognition systems are based on face expression [8], [9] and speech accent [10], [11] where culture specific cues that are common in a group of people originated from the same culture [10] are used. Similarly, just as key words, speech features, or face expression are utilized in these systems, key patterns of people's body movements (i.e. gestures) can be utilized in topic and culture recognition in this research.

III. VARIATIONS IN TOPIC AND CULTURE RECOGNITION

In spite of the topic and culture-specific cues found in the human body movements, these movements are still idiosyncratic of the individual who is performing them [2]. Just as some people are more articulate and do more talking than others, some people make more use of body movements in their interaction than others regardless of their culture and their conversation content.

A. Personality Traits

The term 'personality' is defined by Allport as "the dynamic organization within the individual of those psychophysical systems that determine his unique adjustments to the environment" [12] and "his characteristic behavior and thought" [13]. Allport emphasized on four issues: uniqueness, biology, learning, and consistency [12]. According to Allport, each individual has a unique and consistent set of personality traits that are partly born (i.e. human's biology) and partly inborn (learnt from the culture). For example, after observing videotapes on people performing a variety of tasks, Allport and Vernon [13] found high consistency and uniqueness in handwriting, postures, and gestures over a variety of tasks and situations [13].

Personality traits in this research are unique distributions of poses of communicative body parts [14]. These are distinctive to the individual who is performing them and consistent across situations. The situations in this research are the topics of conversation and the culture of the individual. Although human personality traits are considered as a blessing to many applications as in [15]–[24], these traits are a curse to our topic and culture recognition systems. Thus removing these traits can improve the performance of both systems.

B. Personality Traits: Analysis Methods

One of the manual methods to determine personalities is the questionnaire where the subject chooses the most matched adjectives from a list of adjectives that describes his/her personality (e.g. Briggs and Myer's Big Five [25]) as in [26]. Another manual method is observation, where the researcher watches people performing tasks in videotapes [13] and chooses the most matched adjectives from the Big Five [25] that describes the people's personality.

Traditional person recognition systems are based on speech, where they exploit differences between the distributions of sounds in different speakers, languages [10] [27], and accents [28], [10]. If the accuracy of such systems is very high, then the signals and their distributions are distinctive to the individual who contributed to the systems.

IV. MODELLING BODY MOVEMENT SIGNALS

The recognition systems in this research exploit differences between the distributions of signals in different classes. Classes can be individuals, topics, and cultures. We used the following statistical techniques in building and improving our systems.

A. Gaussian Mixture Models

The core of the three recognition systems is the Gaussian Mixture Models (GMMs). The GMMs are used to model the distributions of human body poses hoping that from this approach we can infer the information needed in the three recognition systems. In other words, we will determine whether or not the similarities of these distributions found in each class; and the differences of these distributions across classes, are sufficient to enable the GMM method to be applied successfully to the relevant recognition problem. The class is one of person, topic, or culture.

B. Class Separation in Pattern Recognition Systems

Each of the GMM based systems above is combined with Support Vector Machines (SVMs) [27] to create a GMM-SVM system. Combining the generative modelling with the discriminative classifier is found to be advantageous in many applications [10], [27]. We hope that the similarities of body poses' distributions in each class and the differences of these distributions between classes in each of our systems above are sufficiently separable to enable the GMM-SVM approach to be applied successfully to our recognition tasks.

C. Variability Compensation Techniques

Inter-Session Variability Compensation (ISVC) [29] and Nuisance Attribute Projection (NAP) [30] are used in speech technology to compensate variations caused by speaker variability [28], channel and environment effects [31]. These techniques will be used to compensate for the individual's personality traits in this research. The variability compensation technique seeks to accommodate all differences between sessions representing a specific class. In topic recognition, sessions are all recorded sessions about a specific topic (e.g. musical instruments) captured from different individuals regardless of their culture. Sessions in culture recognition are those captured from different individuals who originated from the same culture.

The paper is organized as follows. In Section V, we describe the corpus collection experiments. First, we describe three trials conducted prior to our corpus collection. Then we describe the quantity and quality of our corpus. Section VI describes the components of person, culture and topic recognition systems. Section VII presents the results from the GMMs, GMM-SVMs, and the variability compensation techniques. The last section, VIII presents the discussion and the future direction.

V. HUMAN BODY MOVEMENTS CORPUS

To conduct our research, we used a 12-camera system, the Qualisys Track Manager [32] (QTM) to collect three types of

data: 3D motion data, video recording, and analogue speech recording. The three modalities were synchronized together by setting the QTM system as the main system which started and ended the process of data collection from the three modalities. Once triggered, the QTM system sent a signal through synchronisation cables to three pieces of equipment in order to start collecting data from the QTM cameras, the speech equipment, and the video camera. When the session data was collected, the QTM sent a stopping signal to the three of them in order to stop collecting data. To obtain a good quality of data, we ran three trials prior to the corpus collection.

A. First Trial: the QTM and Markers

We studied the quality of data obtained from the QTM system for different numbers of markers attached to the subject's communicative body parts [14]. Twelve cameras were distributed on the ceiling around the QTM laboratory and calibrated as instructed by the QTM user guide. We attached thirteen 12mm-markers to the subject's body and twenty-six 7mm-markers were attached to a pair of gloves using self-adhesive tape (Fig. 1).
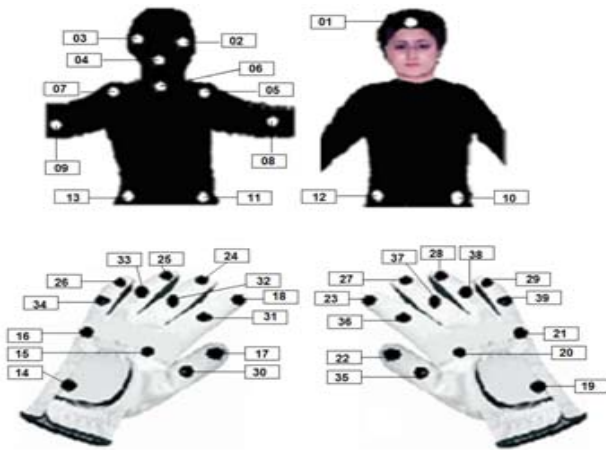

Fig. 1 Markers on the subject's communicative body parts

As the resulting data was not of good quality, we modified our experiment settings. For example, we reduced the number of markers on each finger from two markers to one, brought the QTM cameras closer to the measurement area, and increased the size of the markers.

B. Second Trial: Human Interactions about the Topics

We ran this trial to analyse the utility of the potential topics and the effect of familiarity between a pair of individuals. We video recorded a pair of subjects communicating with each other and asked the subjects to answer questions about their relationship to their partners and their preferred topics.

We found that the most successful topics were musical instruments, wildlife, celebrations and festivals, and routes on a map. In addition, as the degree of familiarity between the subject and his partner affected this trial, we recruited groups of friends to take part in our corpus collection experiments.

C. Third Trial: Piloting Solutions

We ran a third trial to manipulate the solutions proposed in the first and second trials. As tracking the markers on the subject's fingers is still problematic, we captured the movements of thumbs and indexes only as people use these two fingers more frequently than others. However, the quality of 3D motion data, generally, improved considerably compared to the data collected in the first trial.

D. Human Body Movements Corpus

We recruited twenty-seven subjects, arranged them in two groups according to their cultural background, and then further arranged each cultural group into groups of three subjects. In order to maintain consistency throughout, the same experimenter prepared all subjects in the corpus.

Each pair of subjects communicated with each other in their native language and recorded eight 2-minute sessions. We collected another eight 2-minute sessions from each pair of Chinese subjects communicating with each other in English as a non-native language. This gave us a hundred and forty-four 2-minute sessions (72 in English as a native language, 36 in Chinese as a native language, and 36 in English as a non-native language).

In an ideal capture, there would be twenty-three trajectories, each correspond to one of the twenty-three markers attached to the subject's body. The more marker occlusion occurs, the more partial trajectories of each marker will be produced by the QTM system. The corpus is of a good quality as the average rate of broken trajectories is 1.46% in the corpus. This result is a considerable improvement compared to the results in the first and third trials (5.28% and 7.12% respectively).

The training dataset used in all of our experiments consists of 512, 512, and 504 minutes data, as shown in Table I, for topic, culture, and person recognition. We used all data captured from 24 subjects to build culture and topic recognition systems. Whereas, we used 14 and 28 minutes data captured from each British and Chinese respectively to build our person recognition systems.

TABLE I
TRAINING AND EVALUATION DATASETS

Data Sets	Person	Culture	Topic
Training	504min	512min	512min
Evaluation	72min	64min	64min

For evaluating culture and topic recognition systems, we randomly held back all data captured from 2 British and 1 Chinese subject. This is not the case for evaluating person recognition systems as we held back 2 and 4 minutes of data captured from each British and Chinese subject in the corpus respectively. To increase the size of the evaluation data, we divided each 2-minute data into 4 parts each contains 30 seconds.

VI. THE CLASSIFICATION SYSTEMS

We applied three processes to the 3D data signals prior to using them in building our recognition systems. First, we

estimated missing information using the QTM function 'Gap Filling'. Second, as the pair of subjects in each session was facing each other during recording, the orientation of their data is different. Therefore, we mirrored data captured from one of them using a reflection matrix [33] and kept the data captured from the other subject in its original orientation. We plotted the data from both subjects to ensure that data from both subjects appeared on top of each other in the coordinate space. Third, we applied means and variance normalization.

Modeling the distributions of vectors for different classes has been successfully applied in speech research such as in language and speaker recognition systems [28]–[30] using the Gaussian Mixture Models (GMMs). The most common variants of the GMMs are the GMM with the Universal Background Model (GMM-UBM) and the GMM with Support Vector Machines (GMM-SVM). Thus, the GMM approach is the core of our GMM-UBM and GMM-SVM based systems.

A. The GMM Approach

The Universal Background Model (UBM) is built using the training datasets of all classes in the relevant system. Class-dependent models are obtained by MAP adaptation [34], adapting means of the UBM, using the class-specific enrollment data. The result is one UBM and C class-dependent GMMs (where C= 27, 4, and 2 for person, topic, and culture respectively).

We calculated the conditional probability of the input 3D motion signals given some pre-trained motion signals model, typically a GMM. Our calculation is a weighted sum of Gaussian Probability Density Functions (PDF). A GMM, therefore, is a PDF p defined as a linear combination of Gaussian PDFs [35],

$$p(x_t \mid \lambda) = \sum_{i=1}^{M} \frac{w_i}{(2\pi)^{D/2} \mid \sum_i \mid^{1/2}} \exp\left(-\frac{1}{2}(x_t - \mu_i)' \sum_i^{-1}(x_t - \mu_i)\right) \quad (1)$$

where $w_i \geq 0$ and $\sum_{i=1}^{M} w_i = 1$, x_t is a vector with dimension D at the time t, M is the number of Gaussian components in the mixture density and λ is the set of GMM parameters $\{w_i, \mu_i, \Sigma_i\}$ where these three components are the weight, mean, and diagonal covariance of the ith Gaussian component.

For each recognition system and each application we tested the claim that a particular body movement signal belongs to a particular class (person, topic, or culture) by comparing its score rate with the scores of other classes. We assigned the movement signal to the class with the largest score. The percentage accuracy is the percentage of times that this class is correct.

B. The GMM-SVM Approach

Given a session from the training dataset, the GMM-UBM training is performed by MAP adaptation of the means m_i. From this adapted model, we form a GMM supervector. We adapted only GMM means as this method outperforms adapting the means and covariances [10]. The adapted GMM mean vectors are then concatenated into one 'supervector'. The different classes are assumed to be linearly separable in this supervector space.

The SVM training involves learning the parameters of a hyperplane that best separates two classes, in the sense that distances from the hyperplane of the closest supervectors (the margin) is maximized. These closest supervectors are the 'support vectors' that give the SVM its name. The GMM supervector defines a mapping between a session and a high dimensional vector. Then the SVM compares two sessions using a dot product SVM kernel.

We used 'one-against-one' strategy in our culture recognition as we have only two classes and 'one against-the-rest' in person and topic recognition as they are multi-class problems. In 'one against-the-rest' strategy, the supervectors in the relevant system are used to build one SVM for each class by treating that class as the 'target' class and the other classes as the 'background' class.

The test supervectors are scored against the SVM models. Positive scores indicate that the test supervector belongs to the target class while a negative score shows that the test supervector belongs to one of the non-target classes. We used the SVM-KM SVM MATLAB toolbox [36] to train and evaluate our GMM-SVM based systems.

C. The Personality Traits Compensation

The variability due to the individual's personality traits can be dealt with by using a person-dependent topic or culture recognition system, similar to speaker dependent speech recognition [37]. As this method can be expensive, we removed variations in the data that are not relevant to the classification problem. Specifically, in topic and culture recognition, we removed the individual differences.

Inter-Session Variability Compensation (ISVC) [29] and Nuisance Attribute Projection (NAP) [30] are used in speech detection to compensate the variations caused by channel and environment effects [31] and speaker variability [28]. Applying them to the removal of personality traits, we wanted the distortions due to individual's personality traits in the high-dimensional space to be summarized by a small number of parameters in a lower dimensional subspace. Both techniques compensate for session variations (in this case the individual's personality) by removing the estimated influence of these traits.

These techniques can be applied to the GMM model domain or to the feature domain or both of them at the same time [31]. In our research, we applied the NAP to the training supervectors in the model domain and the ISVC to the feature domain of the test segments.

VII. THE RESULTS

The GMM components in each relevant system are trained using all of the training dataset in Table I with 4 EM iterations (chosen empirically) updating all model parameters. First, we investigated the best model order that represent the distribution of the training data.
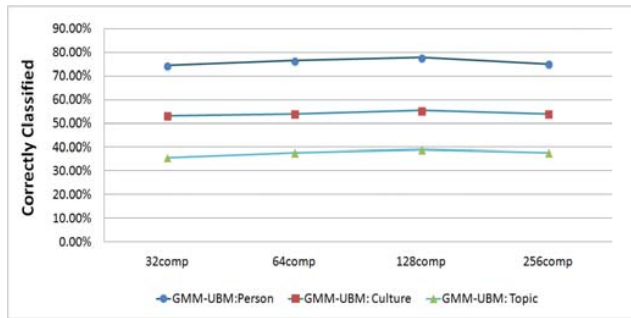
Fig. 2 The Model Order in GMM-UBM

Fig 2 shows that by increasing the number of Gaussian components, the accuracy of the three systems improves up to 128 components and then degrades when 256 components are used suggesting that the amount of training data is insufficient to train a higher order of components. As 128 GMM components show the best performance, we used this order in all GMM based systems.

A. The GMM-UBM versus GMM-SVM

Table II shows that the GMM-SVM systems outperform the GMM-UBM systems in all recognition tasks. We obtained improvement of 85.42%, 62.50%, and 40.63% in the person, topic, and culture recognition tasks respectively. The result suggests that the GMM-SVM systems, being discriminative, are able to focus on the class-specific boundaries.

TABLE II
RESULTS OF GMM-UBM AND GMM-SVM

Methods	Person	Culture	Topic
GMM-UBM	77.78%	55.47%	39.06%
GMM-SVM	85.42%	62.50%	40.63%

Culture recognition is a two- class problem, so assuming equal priors the 'random' performance would be 50%. Hence our GMM-UBM and GMM-SVM based systems are performing slightly better than random. For topic recognition, there are four classes so the performance of a random classifier would be 25% accuracy. Our classifier is doing significantly better so there is information in the signals that is useful for topic classification. These results suggest that differences due to individual's personality traits are much higher than differences due to topic and culture.

A. Variability Compensation Techniques

The dimensions of the 'nuisance' subspaces were 100 and 175 for culture and topic recognition, respectively. Applying these to the removal of personality traits, the idea is that the distortions due to individual's personality traits in the high-dimensional subspace are summarized by a small number of parameters in a lower dimensional subspace.

We compare the accuracy of GMM-SVM based systems where no variability compensation is applied, where NAP is applied to the training supervectors, and where NAP is applied to the training supervectors and the ISVC to test supervectors in the testing phase as shown in Table II, first row in Table III,

and second row in Table III respectively.

TABLE III
RESULTS OF NAP AND ISVC

Methods	Culture	Topic
NAP only	71.88%	42.97%
NAP & ISVC	73.44%	46.09%

We obtained 71.88% and 42.97% from culture and topic recognition respectively when the NAP technique is applied to the training supervectors prior to passing these supervectors to the GMM-SVM for classification. When NAP is applied to the training supervectors and ISVC to the test motion vector, the accuracy of culture and topic recognition systems improved by 73.44% and 46.09% respectively.

The results suggest that many of the individual's personality traits are culture-specific cues that could not be removed from the culture recognition system. Consequently, removing both the individual's personality traits and culture-specific cues, we obtained big improvement in topic recognition systems. The reason of a big improvement for topic and less improvement for culture recognition is that each subject has a unique culture but varying topics. So personality traits might be good for culture recognition.

VIII. CONCLUSIONS AND FUTURE WORK

The main contribution of this paper is to show the viability of recognizing individuals, culture and topic from 3D human motion data. The accuracy of the topic and culture recognition systems is enhanced by removing variations due to personality traits. We hypothesized that removing these variations will improve the accuracy of culture and topic recognition systems. We used person recognition system to obtain evidence that the variations in topic and culture recognition are due to the individual's personality traits. If the accuracy of this system is very high, then the body poses and their distributions are distinctive to the individual who performed these movements. The accuracy of person recognition systems was high (77.78%, 85.42%) proving that the main variations are due to the individual's personality.

The GMM-SVM could separate the classes in culture and topic recognition and consequently the accuracy of the systems improved as in Table III. However, the differences due to the individual's personality traits are much higher than the differences among classes in topic and culture recognition system. The NAP serves the SVM by removing the individual's differences from the training supervectors before introducing these supervectors to the SVM as in first row in Table III. Then we used ISVC to remove the direction of the individual's personality in the testing segments leading to a further improvement, as in second row in Table III, in culture and topic recognition.

The results obtained from topic recognition suggest that this task is a difficult task when it is based on human body movement signals only. Even within a task with homogeneous population of individuals (who originated from the same country), there are likely to be significant variations. For

example, in a certain topic, people may not repeat the same conversation or perform the same gestures in the same manner, speed, or frequency. An extension to this task could be a comparison between human performance and machine performance in topic recognition task.

In this paper, we have 27, 4, and 2 classes in person, topic, and culture recognition respectively. A further investigation is needed after increasing the number of classes in each recognition task. Finally, the result above may raise the possibility of extracting other information such as emotion, mimic behavior, language proficiency, etc. From a broader perspective, the results obtained in this paper raise the possibility of applying variability compensation techniques to the automatic classification of other than person recognition systems such nativity recognition or the effect of non-native language on human body poses' distributions.

REFERENCES

[1] R. A., Bolt, "Put that there: Voice and gesture at the graphics interface", Computer Graphics, vol. 14(3), pp. 262-270, 1980

[2] D. McNeil, "Hand and Mind: What gestures Reveal about Thought", p.37, 1992

[3] D. McNeill, "So You Think Gestures are Nonverbal?", Psychological review, vol. 92 (3), p. 350, 1985

[4] M. Rehm, N. Bee, and E. André, "Wave like an Egyptian: accelerometer based gesture recognition for culture specific interactions", In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, vol. 1, pp. 13-22, 2008.

[5] F. Cavicchio and S. Kita, "English/Italian Bilinguals Switch Gesture Parameters when they Switch Languages" in The Proceedings of Tilburg Gesture Research Meeting, pp. 305-309, 2013

[6] G. W. Allport, "Personality: A psychological Interpretation", New York: Holt, 1937

[7] S. Bird, S. Browning, R. Moore and M. J. Russell, "Dialogue move recognition using topic spotting techniques", Proc. ESCA Workshop on Spoken Dialogue Processing -Theory and Practic, Vigsø, Denmark, pp. 45-48, 1995

[8] B. Dhillon, R. Kocielnik, I. Politis, M. Swerts, and D. Szostak, "Culture and facial expressions: a case study with a speech interface", In Human-Computer Interaction–INTERACT 2011, Springer Berlin Heidelberg, pp. 392-404, 2011.

[9] G. Zen, E. Sangineto, E. Ricci, and N. Sebe, "Unsupervised Domain Adaptation for Personalized Facial Emotion Recognition". In Proceedings of the 16th International Conference on Multimodal Interaction, pp. 128-135, 2014

[10] A. Hanani, M.J. Russell and M. J. Carey, "Human and computer recognition of regional accents and ethnic groups from British English speech". Computer Speech & Language, vol. 27(1), 59-74, 2013.

[11] K. C. Sim, S. Zhao, K. Yu, and H. Liao, "ICMI'12 grand challenge: haptic voice recognition", In Proceedings of the 14th ACM international conference on Multimodal interaction, pp. 363-370, 2012.

[12] G. W. Allport, "Pattern Growth in Personality", New York: Holt, Rinehart, and Winston, p48, 1937.

[13] G. W. Allport and P. E. Vernon, "Studies in Expressive Movement", p28, 1933.

[14] S. D. Kelly, S. M. Manning and S. Rodak, "Gesture Gives a Hand to Language and Learning: Perspectives from Cognitive Neuroscience, Developmental Psychology and Education", Language and Linguistics Compass, vol. 2(4), pp. 569-588, 2008.

[15] L. S. Nguyen, A. Marcos-Ramiro, M. Marrón Romera, and D. Gatica-Perez, D. "Multimodal analysis of body communication cues in employment interviews". In Proceedings of the 15th ACM on International conference on multimodal interaction, pp. 437-444, 2013.

[16] M. Valstar, "Automatic Behaviour Understanding in Medicine". International Conference on Multimodal Interaction (ICMI), 2014

[17] B. De Raad and H. Schouwenburg, "Personality in Learning and Education: A Review", vol. 10, pp. 303-336, 1996.

[18] A. Pena, and A. De Antonio, "Inferring interaction to support collaborative learning in 3D virtual environments through the user's avatar Non-Verbal Communication", International Journal of Technology Enhanced Learning, vol, 2(1), pp. 75-90, 2010.

[19] E. S. Kluft, J. Poteat and R. P. Kluft, "Movement observations in multiple personality disorder: A preliminary report", American Journal of Dance Therapy, vol. 9 (1), pp. 31-46, 2006.

[20] L. Chittaro and M. Serra, "Behavioural Programming of Autonomous Characters Based on Probabilistic Automata and Personality", vol. 15(3-4), pp. 319-326, 2004.

[21] H. Kim, S. S. Kwak, and M. Kim, "Personality design of sociable robots by control of gesture design factors". In Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on, pp. 494-499, 2008

[22] M. S. Nixon, and J. N. Carter, "Automatic recognition by gait", Proceedings of the IEEE, vol. 94(11), pp. 2013-2024, 2006.

[23] J. P. Singh and S. Jain, "Person Identification Based on Gait using Dynamic Body Parameters", In Trendz in Information Sciences & Computing (TISC), pp. 248-252, 2010.

[24] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition", Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14(1), pp. 4-20, 2004

[25] I. B. Myers and P. B. Myers, "Gifts differing: Understanding Personality Type", Nicholas Brealey Publishing, 2010

[26] I. Jraidi, M. Chaouachi and C. Frasson, "A dynamic multimodal approach for assessing learners' interaction experience", In Proceedings of the 15th ACM on International conference on multimodal interaction pp. 271-278, ACM, 2013.

[27] W. M. C. J. P. Campbell, D. A. Reynolds, E. Singer and P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition", Computer Speech & Language, vol. 20 (2), pp. 210-229, 2006

[28] A. Hanani, M. Carey and M. J. Russell, "Improved language recognition using mixture components statistics", In INTERSPEECH, pp. 741-744, 2010.

[29] R. Vogt and S. Sridharan, "Experiments in session variability modelling for speaker verification", In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on vol. 1, pp. I-I, 2006.

[30] W. Campbel, D. Sturim, D. Reynolds and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation". In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on vol. 1, pp. I-I, 2006.

[31] C. Vair, D. Colibro, F. Castaldo, E. Dalmasso, and P. Laface, P. "Channel factors compensation in model and feature domain for speaker recognition", In Speaker and Language Recognition Workshop. IEEE Odyssey, pp. 1-6, 2006.

[32] Qualisys, "Qualisys Track Manager", QTM, 2006

[33] J. C. Gower, "Euclidean Distance Matrices' in 'Convex optimization and Euclidean distance geometry", J. Dattorro, ed., Lulu. Com, 2008

[34] J. Gauvain, and L., Lee, C., "Maximum A-Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains", IEEE Transactions on Speech and Audio Processing vol.2, pp. 291–298, 1994.

[35] C. M. Bishop, "Pattern Recognition and Machine Learning", Springer, New York, p84-87, 2006.

[36] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and Kernel Methods Matlab Toolbox", Perception Systèmes et Information, INSA de Rouen, Rouen, France, http://asi.insa-rouen.fr/enseignants/arakotom/toolbox/, 2005.

[37] R. Vergin, A. Farhat and D. O'Shaughnessy, "Robust gender-dependent acoustic-phonetic modelling in continuous speech recognition based on a new automatic male/female classification", In Spoken Language, ICSLP 96, Proceedings, Fourth International Conference vol. 2, pp. 1081-1084, 1996.