

# Language Independent Speech-driven Facial Animation System

Lucas D. Terissi\* and Juan Carlos Gómez

Laboratory for System Dynamics and Signal Processing  
FCEIA, Universidad Nacional de Rosario  
CIFASIS, CONICET  
Riobamba 245bis, 2000, Rosario, Argentina  
{lterissi, jcgomez}@fceia.unr.edu.ar

**Abstract.** *In this paper, a speech-driven MPEG-4 compliant facial animation system is proposed. The system is based on the inversion of a joint Audio-Visual Hidden Markov Model to estimate the visual information from speech data. The inversion algorithm is derived for the general case of considering full covariance matrices for the audio-visual observations. The proposed speech-driven facial animation system operates at a sub-phonemic level, with the advantage of being adaptable to any language. The experiments are carried out using audio-visual databases compiled by the authors.*

## 1 Introduction

The widespread use of multimedia applications such as computer games, online virtual characters, video telephony, and other interactive human-machine interfaces, has made speech driven animation of virtual characters to play an increasingly important role. Several approaches have been proposed in the literature for speech driven facial animation. Among the different approaches proposed in the literature to model audio-visual data, the ones based on Hidden Markov Models (HMM) have proved to yield more realistic results when used in applications of speech driven facial animation.

Earlier approaches for speech-driven facial animation systems, such as the works in [1], [2], [3] and [4], resort to different HMM structures and require the use of Viterbi optimization algorithm [5] in the training or synthesis stages. This leads to video predictions of limited quality due to the high noise sensitivity of Viterbi algorithm. To address this limitation, Choi *et al* [6] have proposed a Hidden Markov Model Inversion (HMMI) method for audio-visual conversion. HMMI was originally introduced in [7] in the context of robust speech recognition. In HMMI, the visual output is generated directly from the given audio input and the trained HMM by means of an expectation-maximization (EM)

---

\* Author to whom all correspondence should be addressed. This work is partially supported by ANPCyT under grant PICT 11-15132, and by CONICET.

iteration, thus avoiding the use of the Viterbi sequence and improving the performance of the estimation [8]. Recently, Xie *et al* [9] proposed a coupled HMM approach and derived an expectation maximization (EM)-based A/V conversion algorithm for the CHMMs, which converts acoustic speech into decent facial animation parameters.

In this paper, a speech driven MPEG-4 compliant facial animation system is proposed. A joint audio-visual Hidden Markov Model (AV-HMM) is trained using audio-visual data and then Hidden Markov Model inversion is used to estimate the animation parameters from speech data. The feature vector corresponding to the visual information during the training is obtained via Independent Component Analysis (ICA). Previous approaches based on HMMs consider diagonal covariance matrices for the audio-visual observations, invoking reasons of computational complexity. In this paper full covariance matrices are used. This choice is motivated by simulation results obtained in a previous work [10] where the use of full covariance matrices leads to an accurate estimation of the visual parameters, yielding a performance similar to that of using diagonal covariance matrices, but with a less complex model and without affecting significantly the computational load. The proposed model operates at a subphonemic level, mapping speech acoustic features (e.g., cepstral coefficients) onto facial expressions (visual features), avoiding the need for phoneme recognition and phoneme to viseme<sup>1</sup> conversion, making the system adaptable to any language.

The rest of the paper is organized as follows. An overview of the speech driven facial animation system is presented in section 2. The AV-HMM is introduced in section 3, where a Hidden Markov Model Inversion algorithm for the general case of considering full covariance matrices for the audio-visual observations is also derived. In section 4, the proposed algorithm for feature extraction is described. The MPEG-4 compliant facial animation technique is presented in section 5. Experimental results and some concluding remarks are included in sections 6 and 7, respectively.

## 2 Speech Driven Facial Animation System Overview

A block diagram of the proposed speech driven animation system is depicted in Fig. 1. An audiovisual database is used to estimate the parameters of a joint AV-HMM. This database consists of videos of a talking person with reference marks in the region around the mouth, see Fig. 2(a).

In a first training stage, feature parameters of the audiovisual data are extracted. The audio part of the feature vector consists of mel-cepstral coefficients, while the visual part are the coefficients in a ICA representation of the above mentioned set of reference marks. In a second training stage, the audio part of the AV-HMM is re-trained using audio data from a speech-only database. Re-training only the audio part of the model allows to obtain a more robust model against inter-speaker variability, avoiding the need to record videos of speakers with the reference marks on their faces.

<sup>1</sup> A viseme is defined as a basic unit of speech in the visual domain.

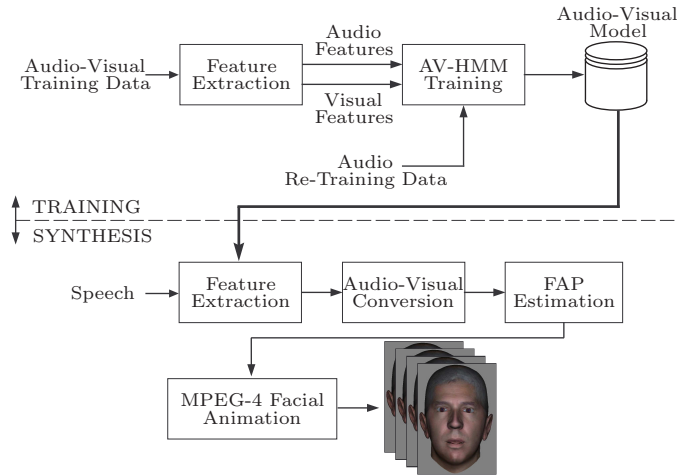


Fig. 1. Schematic representation of the speech driven animation system.

For the speech driven animation, speech data is used to estimate the visual features by inversion of the AV-HMM using a technique described in section 3. From these data, Facial Animation Parameters (FAPs) of the MPEG-4 [11] standard are computed to generate the facial animation.

### 3 Audio Visual Model

In this paper, a joint AV-HMM is used to represent the correlation between the speech and facial movements. The AV-HMM, denoted as  $\lambda_{av}$ , is characterized by three probability measures, namely, the state transition probability distribution matrix ( $A$ ), the observation symbol probability distribution ( $B$ ) and the initial state distribution ( $\pi$ ), and a set of  $N$  states  $S = (s_1, s_2, \dots, s_N)$ , and audiovisual observation sequence  $O_{av} = \{o_{av1}, \dots, o_{avT}\}$ . In addition, the observation symbol probability distribution at state  $j$  and time  $t$ ,  $b_j(o_{avt})$ , is considered a continuous distribution which is represented by a mixture of  $M$  Gaussian distributions

$$b_j(o_{avt}) = \sum_{m=1}^M c_{jm} \mathcal{N}(o_{at}, o_{vt}, \mu_{jm}, \Sigma_{jm}) \quad (1)$$

where  $c_{jm}$  is the mixture coefficient for the  $m$ -th mixture at state  $j$  and  $\mathcal{N}(o_{at}, o_{vt}, \mu_{jm}, \Sigma_{jm})$  is a Gaussian density with mean  $\mu_{jm}$  and covariance  $\Sigma_{jm}$ . The audiovisual observation  $o_{avt}$  is partitioned as  $o_{avt} \triangleq [o_{at}^T, o_{vt}^T]^T$ , where  $o_{at}$  and  $o_{vt}$  are the audio and visual observation vectors, respectively.

A single ergodic<sup>2</sup> HMM is proposed to represent the audiovisual data. An alternative to an ergodic model, would be a set of left-to-right HMMs representing the different phonemes (with associated visemes) of the particular language. These models have been used in the context of speech modeling by several authors, see for instance [9]. An ergodic model provides a more compact representation of the audiovisual data, without the need of phoneme segmentation, which is required when left-to-right models are used. In addition, this has the advantage of making the system adaptable to any language.

### 3.1 AV-HMM Training

The training of the AV-HMM consists of two stages, each one using a different database. In the first training stage, an audiovisual database consisting of a set of videos of a single talking person with reference marks drawn on the region around the mouth, is used to estimate the parameters of an ergodic AV-HMM, resorting to the standard Baum-Welch algorithm [12]. Details on the composition of the audiovisual feature vector are given in Section 4, where procedures to take into account audio-visual synchronization and co-articulation are also described. In the second training stage, a speech-only database consisting of audio recordings from a set of talking persons is employed to re-train the audio part of the AV-HMM, leading to a speaker independent model. The re-training is carried out using an only audio HMM (hereafter denoted as A-HMM), with the same structure, which is constructed from the AV-HMM. The A-HMM has the same transition probability and initial state probability matrices obtained in the first stage, while the corresponding observation symbol probability distribution is re-estimated from the speech-only database. The observation symbol probability distribution is parameterized by  $\mu_{jm}$ ,  $\Sigma_{jm}$  and  $c_{jm}$ , see equation (1). To emphasize the mix composition of the AV-HMM, the mean and covariance parameters can be partitioned as

$$\mu_{jm} = \begin{bmatrix} \mu_{jm}^a \\ \mu_{jm}^v \end{bmatrix} \quad \Sigma_{jm} = \begin{bmatrix} \Sigma_{jm}^a & \Sigma_{jm}^{av} \\ \Sigma_{jm}^{va} & \Sigma_{jm}^v \end{bmatrix} \quad (2)$$

where the superscript  $a$  and  $v$  denote the audio and visual parts, respectively. During the second training stage, only  $\mu_{jm}^a$  and  $\Sigma_{jm}^a$  are re-estimated using the speech-only data. Finally, the re-estimated parameters are fed back into the AV-HMM.

### 3.2 Audio-to-Visual Conversion

Hidden Markov Model Inversion (HMMI) was originally proposed in [7] in the context of robust speech recognition. Choi and co-authors [6] used this technique to estimate the visual features associated to audio features for the purposes of speech driven facial animation. Typically, it is assumed [7], [6], [9] a diagonal

<sup>2</sup> An ergodic model is one in which transitions among all the states are allowed.

structure for the covariance matrices of the Gaussian mixtures, invoking reasons of computational complexity. This assumption is relaxed in this paper allowing for full covariance matrices. This leads to more general expressions for the visual feature estimates.

The idea of HMMI for audio-to-visual conversion is to estimate the visual features based on the trained AV-HMM, in such a way that the probability that the whole audiovisual observation has been generated by the model is maximized. It has been proved [12] that this optimization problem is equivalent to the maximization of the auxiliary function

$$\begin{aligned}
 Q(\lambda_{av}; \lambda_{av}, O_a, O_v, O'_v) &\triangleq \\
 &\triangleq \sum_{j=1}^N \sum_{m=1}^M P(O_a, O_v, j, m | \lambda_{av}) \log P(O_a, O'_v, j, m | \lambda_{av}) \\
 &= \sum_{j=1}^N \sum_{m=1}^M P(O_a, O_v, j, m | \lambda_{av}) \left[ \log \pi_{j_0} + \sum_{t=1}^T \log a_{j_{t-1}j_t} + \right. \\
 &\left. + \sum_{t=1}^T \log \mathcal{N}(o_{at}, o'_{vt}, \mu_{j_t m_t}, \Sigma_{j_t m_t}) + \sum_{t=1}^T \log c_{j_t m_t} \right] \tag{3}
 \end{aligned}$$

that is

$$O'_v = \arg \max_{O'_v} \{Q(\lambda_{av}; \lambda_{av}, O_a, O_v, O'_v)\} \tag{4}$$

where  $O_a$ ,  $O_v$  and  $O'_v$  denote the matrices containing the audio, visual and estimated visual sequences from  $t = 1, \dots, T$ , respectively,  $\pi_{j_0}$  denotes the initial probability for state  $j$  and  $a_{j_{t-1}j_t}$  denotes the state transition probability from state  $j_{t-1}$  to state  $j_t$ .

The solution to the optimization problem in (4) can be computed by equating to zero the derivative of  $Q$  with respect to  $o'_{vt}$ . Considering that the only term that depends on  $o'_{vt}$  is the one involving the Gaussians, this derivative can be written as

$$\begin{aligned}
 \frac{\partial Q(\lambda_{av}; \lambda_{av}, O_a, O_v, O'_v)}{\partial o'_{vt}} &= \sum_{j=1}^N \sum_{m=1}^M P(O_a, O_v, j, m | \lambda_{av}) \times \\
 &\times \frac{\partial}{\partial o'_{vt}} \left[ \sum_{t=1}^T \log \mathcal{N}(o_{at}, o'_{vt}, \mu_{j_t m_t}, \Sigma_{j_t m_t}) \right] = 0 \tag{5}
 \end{aligned}$$

Considering that

$$\begin{aligned}
 \log \mathcal{N}(o_{at}, o'_{vt}, \mu_{j_t m_t}, \Sigma_{j_t m_t}) &= \log \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma_{j_t m_t}|}} - \\
 &- \frac{1}{2} \begin{bmatrix} o_{at} - \mu_{j_t m_t}^a \\ o_{vt} - \mu_{j_t m_t}^v \end{bmatrix}^T \begin{bmatrix} \Phi_{j_t m_t}^a & \Phi_{j_t m_t}^{av} \\ \Phi_{j_t m_t}^{va} & \Phi_{j_t m_t}^v \end{bmatrix} \begin{bmatrix} o_{at} - \mu_{j_t m_t}^a \\ o_{vt} - \mu_{j_t m_t}^v \end{bmatrix} \tag{6}
 \end{aligned}$$

where  $d$  is the dimension of  $o_{avt}$  and

$$\Sigma_{j_t m_t}^{-1} = \begin{bmatrix} \Phi_{j_t m_t}^a & \Phi_{j_t m_t}^{av} \\ \Phi_{j_t m_t}^{va} & \Phi_{j_t m_t}^v \end{bmatrix},$$

the estimated visual observation becomes

$$o'_{vt} = \left[ \sum_{j=1}^N \sum_{m=1}^M P(o_a, o_v, j, m \mid \lambda_{av}) \Phi_{jm}^v \right]^{-1} \times \sum_{j=1}^N \sum_{m=1}^M P(o_a, o_v, j, m \mid \lambda_{av}) [\Phi_{jm}^v \mu_{jm}^v - \Phi_{jm}^{va} (o_{at} - \mu_{jm}^a)] \quad (7)$$

For the case of diagonal matrices, equation (7) reduces to

$$o'_{vt} = \left[ \sum_{j=1}^N \sum_{m=1}^M P(o_a, o_v, j, m \mid \lambda_{av}) \Phi_{jm}^v \right]^{-1} \times \sum_{j=1}^N \sum_{m=1}^M P(o_a, o_v, j, m \mid \lambda_{av}) \Phi_{jm}^v \mu_{jm}^v \quad (8)$$

which is equivalent to the equation derived in [6].

As is common in HMM training, the estimation algorithms (7) and (8) are implemented in a recursive way, initializing the visual observation randomly.

### 4 Feature Extraction

The audio signal is partitioned in frames with the same rate as the video frame rate. A number of Mel-Cepstral Coefficients in each frame ( $a_t$ ) are used in the audio part of the feature vector. To take into account the audiovisual co-articulation, several frames are used to form the audio feature vector  $o_{at} = [a_{t-t_c}^T, \dots, a_{t-1}^T, a_t^T, a_{t+1}^T, \dots, a_{t+t_c}^T]^T$  corresponding to the visual feature vector  $o_{vt}$ .

For the visual part, the coefficients in an Independent Component representation of the coordinates of marks in the region around the mouth of the speaking person are used, see Fig. 2(a).

Let  $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T\}$  represent the training data collected from videos. Each vector  $\mathbf{f}_t = [x_1^{(t)}, x_2^{(t)}, \dots, x_P^{(t)}, y_1^{(t)}, y_2^{(t)}, \dots, y_P^{(t)}]^T$  contains the coordinates  $(x_p^{(t)}, y_p^{(t)})$  of each mark ( $p = 1, 2, \dots, P$ ) for the  $t$ -th frame,  $t = 1, 2, \dots, T$ .

Let  $\mathbf{f}_0$  be the neutral facial expression, mainly defined as the expression with all face muscles relaxed and the mouth closed [11]. The relative facial deformation (with respect to the neutral expression) at each frame can be computed as  $\mathbf{d}_t = \mathbf{f}_t - \mathbf{f}_0$ , and a deformation matrix can then be defined as

$$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_T] \quad (9)$$

The different facial expressions in the training data are represented by the columns of matrix  $\mathbf{D}$ . The idea is to represent any facial expression as the linear combination of a reduced number of independent vectors. The dimensionality reduction can be performed by Principal Component Analysis [13]. The PCA stage yields an uncorrelated set of vectors. It is desirable to have a statistically independent set of vector so that information contained in each vector will not provide information on any of the others. This is the main idea in ICA. Summarizing, ICA after PCA will be performed on the data matrix  $\mathbf{D}$ .

Several algorithms are available in the literature for ICA computation. The reader is referred to [13] and the references therein. In this paper, the symmetric decorrelation based FastICA algorithm as implemented in [14] was employed.

As a result of the ICA processing, any facial deformation can then be computed as

$$\mathbf{f}_t = \sum_{k=1}^K o_{vt_k} \mathbf{u}_k + \mathbf{f}_0 \quad (10)$$

where  $\{\mathbf{u}_k\}_{k=1}^K$  are the independent components from  $\mathbf{D}$  and  $o_{vt_k}$  is the  $k$ -th component of the visual vector  $o_{vt}$ . The coefficients  $o_{vt_k}$  are computed in two stages. In the first stage, the mark locations are estimated using image processing techniques. In the second stage, the coefficients  $o_{vt_k}$  are computed in such a way that the facial expression is given by the linear combination of the ICs vectors that best match the mark estimation computed in the first stage. Details of this procedure can be found in [15].

## 5 Facial Animation

As already mentioned, the facial animation technique proposed in this paper is MPEG-4 compliant. The MPEG-4 standard defines 64 Facial Animation Parameters and 84 Feature Points (FPs) on a face model in its neutral state [16]. FAPs represent a complete set of basic facial actions such as head motion, and eye, cheeks and mouth control. FPs are used as reference points to perform the facial deformation.

Based on the estimated facial expression for each frame, the associated FAPs can be determined by computing the displacement of a set of marks from their corresponding position in the neutral facial expression. For instance, the marks encircled in red in Fig. 2(a) can be associated to FAP3 corresponding to jaw opening. Figure 2(b) shows the resulting expression after applying the estimated FAP3 to the neutral expression (several other FAPs, in addition to FAP3, have also been applied to produce the mouth opening and cheek movements). Similarly, several subsets of marks can be associated to the different FAPs.

The avatar's head model, shown in Fig. 2(b), consists of a triangular mesh with 5860 vertices and 11492 faces, provided with separate models for tongue, teeth and eyes. The model, created following the technique described in [17], has an associated texture to provide it with a more realistic look. From this model an MPEG-4 compliant animatable model is generated using the open source software XFaceEd by Balcy [18].



**Fig. 2.** (a) Real person facial expression. Marks associated to FAP3 are encircled in red. (b) Synthesized facial expression.

## 6 Experimental Results

For the audio-visual training, videos of a talking person with reference marks on the region around the person's mouth were recorded at a rate of 30 frames per seconds, with a resolution of  $(320 \times 240)$  pixels. The audio was recorded at 11025Hz synchronized with the video. The videos consist of sequences of the utterances corresponding to 26 Spanish sentences, where each sentence is pronounced 4 times. For the re-training of the audio part of the AV-HMM, an only-audio database consisting of recordings of sequences of the utterances corresponding to 26 Spanish sentences (the same sentences 26 used in the audio-visual training) by 8 speakers (balance proportion of males and females) was collected, where each speaker pronounced each sentence 4 times. For both audio-visual training and audio re-training, the training sentences were arranged in random order 10 times.

Experiments were performed with AV-HMM with full covariance matrices, different number of states and mixtures in the ranges  $[8, 30]$  and  $[1, 5]$ , respectively, and different values of the co-articulation parameter  $t_c$  in the range  $[0, 5]$ . In the experiments, the audio feature vector  $a_t$  is composed by the first eleven non-DC Mel-Cepstral coefficients, while the visual feature vector  $o_v$  is of dimension two ( $K = 2$  in equation (10)). The performances of the different models were quantified by computing the Average Mean Square Error (AMSE)( $\epsilon$ ), and the Average Correlation Coefficient (ACC)( $\rho$ ) between the true and estimated visual parameters, defined as

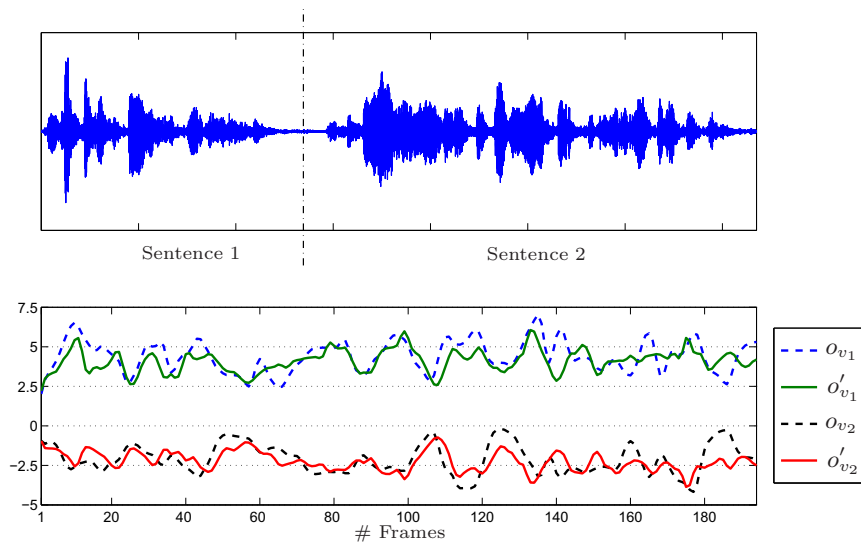
$$\epsilon = \frac{1}{TK} \sum_{k=1}^K \frac{1}{\sigma_{v_k}^2} \sum_{t=1}^T [o'_{vt_k} - o_{vt_k}]^2 \quad (11)$$

$$\rho = \frac{1}{TK} \sum_{t=1}^T \sum_{k=1}^K \frac{(o_{vt_k} - \mu_{v_k})(o'_{vt_k} - \mu'_{v_k})}{\sigma_{v_k} \sigma'_{v_k}} \quad (12)$$

respectively, where  $\mu_{v_k}$  and  $\sigma_{v_k}$  denote the mean and the variance of the true visual observation, respectively, and  $\mu'_{v_k}$  and  $\sigma'_{v_k}$  denote the mean and variance

of the estimated visual parameters, respectively. For the quantification of the visual estimation accuracy, a separate audio-visual dataset, different from the training dataset, was employed.

The true and estimated visual parameters for the case of full covariance matrices with  $N = 20$  states,  $M = 2$  mixtures and co-articulation parameter  $t_c = 3$  are represented in Fig. 3, where a good agreement can be observed. The AMSE and the ACC are for this case  $\epsilon = 0.6325$  and  $\rho = 0.6442$ , respectively. This combination of number of states, number of mixtures and co-articulation parameter was among the ones that yield better results.



**Fig. 3.** Top: Speech signal for two sentences. Bottom: True (dashed line) and estimated (solid line) visual observations associated to the two sentences.

The estimated visual parameters were converted to the associated FAPs in order to produce the speech-driven animation. A good synchronization between the avatar and the actual person was observed. These results are not included here due to space limitations. A video with the animation can be downloaded from <http://www.fceia.unr.edu.ar/lzd/mrg>.

## 7 Conclusions

A speech driven MPEG-4 compliant facial animation system was introduced in this paper. A joint AV-HMM is proposed to represent the audio-visual data and an algorithm for HMM inversion was derived for the general case of considering full covariance matrices for the audio-visual observations. The proposed speech-driven facial animation system operates at a subphonemic level, with the

advantage of being adaptable to any language. The experiments were carried out using audio-visual databases compiled by the authors. Simulation results show a good performance of the proposed algorithms.

## References

1. Yamamoto, E., Nakamura, S., Shikano, K.: Lip movement synthesis from speech based on Hidden Markov Models. *Speech Communication* **26**(1-2) (1998) 105–115
2. Rao, R., Chen, T., Mersereau, R.: Audio-to-visual conversion for multimedia communication. *IEEE Transactions on Industrial Electronics* **45**(1) (Feb 1998) 15–22
3. Chen, T.: Audiovisual speech processing. *IEEE Signal Processing Magazine* **18**(1) (January 2001) 9–21
4. Brand, M.: Voice puppetry. In: *Proceedings of SIGGRAPH, Los Angeles, CA USA* (August 1999) 21–28
5. Viterbi, A.J.: Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Transactions on Information Theories* **13** (April 1967) 260–269
6. Choi, K., Luo, Y., Hwang, J.: Hidden Markov Model inversion for audio-to-visual conversion in an MPEG-4 facial animation system. *Journal of VLSI Signal Processing* **29**(1-2) (2001) 51–61
7. Moon, S., Hwang, J.: Noisy speech recognition using robust inversion of Hidden Markov Models. In: *Proceedings of IEEE International Conf. Acoust., Speech, Signal Processing*. (1995) 145–148
8. Fu, S., Gutierrez-Osuna, R., Esposito, A., Kakumanu, P., Garcia, O.: Audio/visual mapping with cross-modal Hidden Markov Models. *IEEE Transactions on Multimedia* **7**(2) (April 2005) 243–252
9. Xie, L., Liu, Z.Q.: A coupled HMM approach to video-realistic speech animation. *Pattern Recognition* **40** (2007) 2325–2340
10. Terissi, L.D., Gómez, J.C.: Audio-to-visual conversion via HMM inversion for speech-driven facial animation. To appear in *Lecture Notes in Artificial Intelligence Series*, Springer-Verlag (2008)
11. ISO/IEC IS 14496-2, Visual: (1999)
12. Baum, L.E., Sell, G.R.: Growth functions for transformations on manifolds. *Pacific Journal of Mathematics* **27**(2) (1968) 211–227
13. Hyvärinen, A., Karhunen, J., Oja, E.: *Independent Component Analysis*. John Wiley & Sons, Inc., New York (2001)
14. Gävert, H., Hurri, J., Särelä, J., Hyvärinen, A.: *FastICA package for MATLAB*. Laboratory of Computer and Information Science, Helsinki University of Technology (2005)
15. Terissi, L.D., Gómez, J.C.: Facial motion tracking and animation: An ICA-based approach. In: *Proceedings of 15th European Signal Processing Conference, Poznań, Poland* (September 3-7 2007) 292–296
16. Ostermann, J.: Face Animation in MPEG-4. In: *MPEG-4 Facial Animation - The Standard, Implementation and Applications*. John Wiley & Sons (2002) 17–56
17. Ferradal, S., Gómez, J.C.: Generating textures for 3D head models: A wavelet-based approach. In: *Proceedings of ICIP 2006, Atlanta, GA, IEEE* (October 2006)
18. Balcy, K.: XfaceEd: Authoring tool for embodied conversational agent. In: *Proceedings of the International Conference on Multimodal Interfaces 2005, Trento, Italy* (4-6 October 2005) 208–213