Using Dynamic Bayesian Networks for Posed versus Spontaneous Facial Expression Recognition

Melinda Seckington

September 2011

Abstract

Automatic analysis of facial expressions is a complex area of pattern recognition and computer vision with many unresolved problems, one of which is the distinction between posed and spontaneous expressions of emotions. Previous psychology research indicates that the temporal dynamics in the face are essential for distinguishing between posed and spontaneous smiles. There are six temporal characteristics which are important: morphology, apex overlap, symmetry, total duration, speed of onset and speed of offset. In this work, we propose to distinguish between posed and spontaneous expressions by using Dynamic Bayesian networks (DBN) to model the temporal dynamics. The DBN provides a suitable framework to represent probabilistic relationships between and within the various types of temporal dynamics. Based on the temporal phases of four different Action Units (onset, apex offset and neutral of facial actions) and the six temporal characteristics from the psychology research, we build several DBN models to distinguish between posed and spontaneous expressions. We present experimental results from 50 videos displaying posed and spontaneous smiles. When the DBNs trained on the temporal characteristics are combined to provide a joint classification, we attain an AUC of 0.97.

1 Introduction

Traditionally the field of human-computer interaction (HCI) deals with the study of interaction between people and computers. Currently, it is highly insensitive to the affective state of a person, depending instead on passive instruments such as mouse and keyboard. While this may be sufficient for current applications and tasks with computers, the next-generation of HCI designs must take a step further and be able to detect, understand and respond to the various states of a person. The interaction between humans and computers should be as natural as the communication between humans and other humans. Facial expressions are a key element of non-verbal communication between humans, and the automatic analysis of facial expressions is a challenging area in computer vision and pattern recognition.

Many of the existing facial expression analyzers developed so far attempt to recognize a set of six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise) [12] and it is aimed at the analysis of posed data, with test subjects deliberately producing an explicit emotional facial expression. It is only recently that researchers have begun concentrating on spontaneous facial expression data [1] [3], and on the analysis of posed versus spontaneous facial expressions [15] [4].

Research within the psychology field into the differentiation between posed and spontaneous facial expressions indicates that the temporal dynamics of certain facial muscles are very important [7] [6]. For instance, it has been shown that for posed and spontaneous smiles temporal and dynamic characteristics, like duration, co-occurrence and speed, are essential in distinguishing between the two classes [5]. Some of the past work in the field have used temporal dynamics of facial expression in combination with support vector machines [15] and linear discriminant classifiers [4].

A Dynamic Bayesian network (DBN) is a probabilistic graphical model consisting of probabilistic relationships among sets of variables. It is capable of representing the relationships between and within the temporal dynamics of facial muscle movements, and it provides known inference and parameter learning techniques. As such, the various temporal dynamics of facial expressions can be well modelled with a DBN.

This paper reports on our method for posed and spontaneous facial expression recognition by using Dynamic Bayesian networks (DBN) to model the temporal cues. We focus on several temporal dynamics suggested by Ekman in [7]: morphology, apex overlap, symmetry, total duration, speed of onset and speed of offset. Based on these characteristics we build several DBN models to distinguish between posed and spontaneous expressions. The aim of this study is not to evaluate the performance of a fully automated system to distinguish between posed and spontaneous facial expressions, but to investigate whether the temporal dynamics are, as psychologists claim, to be important to posed versus spontaneous recognition and whether they can be modelled in a DBN. For that reason, we use the ground truth manually annotated data as input for our DBN model.

The rest of this paper is outlined as follows. Section 2 explains the Dynamic Bayesian network theory. Section 3 describes the facial expression features used based on the temporal characteristics from the psychology research. Section 4 presents the DBN models. Section 5 describes the

dataset used. Section 6 provides the experimental results. Finally, section 7 analyzes the conclusions drawn from the study.

2 Dynamic Bayesian Networks

A Dynamic Bayesian Network (DBN) is a probabilistic graphical model that can encode the full joint probability distribution for a set of variables [14]. It is a directed acyclic graph (DAG), where each node represents a random variable, and where each arc (also called edges or links) represents the conditional dependency among the variables. Arcs exist between nodes that are dependent on each other; nodes which are not connected indicate that those variables are conditionally independent of each other. The DBN can be seen as a series of time slices; each time slice contains nodes and arcs that describe the domain at a specific moment, and between each time slice arcs exist between nodes to describe the relationship of time.

To be exact, a Dynamic Bayesian Network B is defined by $D = (G, \Theta)$ with the following elements:

1. the directed acyclic graph $G = (V, A)$, where
   - V is a non-empty and finite set of nodes $V = \{\mathbf{X}_1, ..., \mathbf{X}_t\}$ with $\mathbf{X} = \{X^1, ..., X^n\}$. $\mathbf{X}_t$ is used to denote the set of variables $\{X^1, ..., X^n\}$ at time slice t, while X^i represents a single random variable.
   - $A \subseteq V \times V$ is the set of directed arcs between nodes. The *intra-slice* arcs define the relationships between nodes within a time slice, while the *inter-slice* arcs define the relationships between nodes between time slices.
   - if there is a directed arc from node X^i to node X^j, X^i is called a ***parent*** of X^j, while X^j is called a ***descendent*** or ***child*** of X^i. $Par(X^i)$ denotes the set of all parents for X^i, while $Par(\mathbf{X}_t)$ denotes the set of all parents within $\mathbf{X}_t$.
2. the set of conditional probability distributions Θ, that indicate the dependency between nodes:
   - $\mathbf{P}(\mathbf{X}_0)$: the **prior distribution** over the state variables
   - $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$: the **transition model**, which specifies the conditional probabilities of the inter-slice relationships.
   - $\mathbf{P}(\mathbf{X}_t|Par(\mathbf{X}_t))$: the **sensor model**, which specifies the conditional probabilities of the intra-slice relationships.

The three distributions give us a specification of the complete joint distribution over all the variables. For any finite t, the complete joint distribution for a first-order Markov process is:

$$P(\mathbf{X}_0, ..., \mathbf{X}_t) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i|\mathbf{X}_{i-1}) P(\mathbf{X}_i|Par(\mathbf{X}_i)) \quad (1)$$

The transition and sensor models are assumed to be stationary; although the variables change over time, the parameters G and Θ governing these variables do not. Because of this, only the models for the first slice need to be specified. By copying the first slice, the complete DBN can be constructed.

Since the DBN provides the complete joint distribution, each entry in the joint distribution can be calculated from the information in the network.

3 Facial Expression Features

In 1976 Ekman and Friesen developed the Facial Action Coding System (FACS), a comprehensive method to describe all possible visually distinguishable facial movements. The system defines a collection of rules for 32 Action Units (AUs), each indicative of the smallest visually discernible independent facial muscle movement. An AU has 4 temporal phases: the neutral phase, the onset phase, the apex phase, and the offset phase. In figure 1 the flow of these phases is depicted: a facial muscle movement can turn from an onset phase into an apex phase or an offset phase, but can never directly move into a neutral phase. Typically a basic expression holds to the pattern neutral to onset to apex to offset and back to neutral, but more complex expressions can have multiple onsets, apexes or offset (for instance, neutral to onset to apex to onset to apex to offset to neutral).

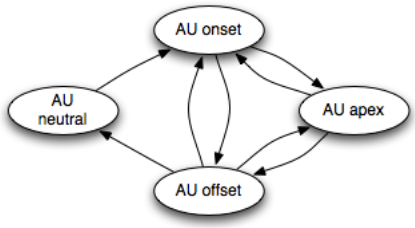

Figure 1: Flow of the temporal phases of an action unit

There are several temporal dynamics based on AUs and their temporal phases that can help distinguish spontaneous facial expressions from posed facial expressions. We turn each temporal characteristic into boolean features to be used in the DBN models.

Morphology: The morphology refers to the occurrence of an AU and its temporal phases. Named after the 19th century physician Duchenne de Boulogne, who discovered it, a Duchenne smile occurs when there are contractions of both the zygomatic major (AU12 - which raises the corners of the mouth) and the orbicularis oculi (AU 6/7 - which raises the

cheek and tightens the upper and lower eye lid). A non-Duchenne smile only involves the zygomatic major, and no involvement of the orbicularis oculi. Ekman reported that the absence of the orbicularis oculi is a strong indicator of a posed smile; however, the presence of the orbicularis oculi does not necessarily mean a spontaneous smile [7].

We define four features for each of the temporal phases:

$$AU_{onset} = \{0, 1\} \quad (2)$$
$$AU_{apex} = \{0, 1\} \quad (3)$$
$$AU_{offset} = \{0, 1\} \quad (4)$$
$$AU_{neutral} = \{0, 1\} \quad (5)$$

These four phases are mutually and collectively exclusive; at any given time, only one of these phases must be active.

Apex Overlap: Ekman reported in [5] that in spontaneous expressions in which there are multiple independent facial actions, it is likely that the apexes of these actions will overlap.

For an AU combination, an apex overlap occurs in a frame when both AUs are in the apex phase:

$$ApexOverlap(AU^X, AU^Y) = \begin{cases} 1 & if\ (AU^X_{apex} = 1) \cap (AU^Y_{apex} = 1) \\ 0 & otherwise \end{cases} \quad (6)$$

Asymmetry: Ekman, Hager and Friesen reported that asymmetries were more frequent in posed smiles than in spontaneous smiles [8]. When asymmetries occurred in posed smiles, they were usually stronger on the left side of the face. In the cases that asymmetries did occur during spontaneous smiles, the asymmetries were equally divided between those stronger on the left and right sides of the face.

For the right mouth corner R, we define $R_{x,t}$ and $R_{y,t}$ as the coordinates of the x- and y- directions at time t. Similarly, for the the left mouth corner L we define $L_{x,t}$ and $L_{y,t}$ as the coordinates of the x- and y- directions at time t. For each frame, asymmetry is determined by first calculating the difference of displacement d of the left and right mouth corners. This is normalized by the overall displacement of the left and right mouth corners.

$$d_{right} = \sqrt{(R_{x,t} - R_{x,t-1}) - (R_{y,t} - R_{y,t-1})} \quad (7)$$
$$d_{left} = \sqrt{(L_{x,t} - L_{x,t-1}) - (L_{y,t} - L_{y,t-1})} \quad (8)$$
$$a = \frac{abs(d_{right} - d_{left})}{abs(d_{right} + d_{left})/2} \quad (9)$$

We define a threshold α to distinguish between symmetry and asymmetry. For each video, we obtain a symmetry feature:

$$Symmetry = \begin{cases} 1 & if\ a > \alpha \\ 0 & otherwise \end{cases} \quad (10)$$

Total duration: Ekman and Friesen observed in [7] that most spontaneous smiles (using AU 12 as an indicator to a smile) were between 2/3s of a second and 4 seconds, while posed false smiles were likely to last longer. Hess and Kleck [9] confirm that posed expressions are longer than spontaneous expressions, experimenting with facial expressions of happiness and disgust.

For an AU, we define the total duration as the number of seconds an AU is active, i.e. whether an AU is in the onset, apex or offset phase. As the temporal phases are mutually and collectively exclusive, this can also be defined as when an AU is not in the neutral phase:

$$td = length(AU_{neutral} = 0) \quad (11)$$

We define a threshold β to distinguish between "short" and "long" durations. For each video, we obtain a total duration feature:

$$TotalDuration = \begin{cases} 1 & if\ td > \beta \\ 0 & otherwise \end{cases} \quad (12)$$

Speed: Ekman suggested in [5] that the onset of a posed expression will be often more abrupt than that of a spontaneous expression. Hess and Kleck confirmed this and reported that in comparison to spontaneous smiles, posed smiles are quicker in onset and offset time [9].

For each video, the speed of onset of a smile is defined by the displacement of the mouth corners divided by the duration of the smile. We only look here at the right mouth corner, using the definition of R from above.

$$so = \frac{\sqrt{(R_{x,t} - R_{x,1}) - (R_{y,t} - R_{y,1})}}{length(AU_{onset} = 1)} \quad (13)$$

We define a threshold γ to distinguish between "quick" and "long" speed of onset. For each video, we obtain a speed of onset feature:

$$SpeedOnset = \begin{cases} 1 & if\ so > \gamma \\ 0 & otherwise \end{cases} \quad (14)$$

We define a similar equations for the speed of offset, but with a threshold δ:

$$sf = \frac{\sqrt{(R_{x,t} - R_{x,1}) - (R_{y,t} - R_{y,1})}}{length(AU_{offset} = 1)} \quad (15)$$
$$SpeedOffset = \begin{cases} 1 & if\ sf > \delta \\ 0 & otherwise \end{cases} \quad (16)$$

4 The DBN Models

In this study, several DBN models were made based on the temporal characteristics from the psychology research. One of the goals of this paper was to investigate how each temporal characteristic could be modelled in a DBN, and whether it would contribute to a better classification.

All DBN models were created with the Bayes Net Toolbox for Matlab [10], which supports different types of probability distributions, exact and approximate inference, parameter and structure learning, and static and dynamic models. The models created here all use a dynamic structure, boolean nodes and the Junction tree algorithm for inference [11].

4.1 Temporal Phases and Morphology

This DBN models the temporal phases of Action Units and can be seen in figure 2. We use the graphical notation of *plates* from Bishop [2] to indicate duplicate sets of nodes. Depending on the number of AUs we want to model, the DBN consists of $1 + (4 * N)$ nodes. For each time slice, we define the set $\mathbf{X}_t$:

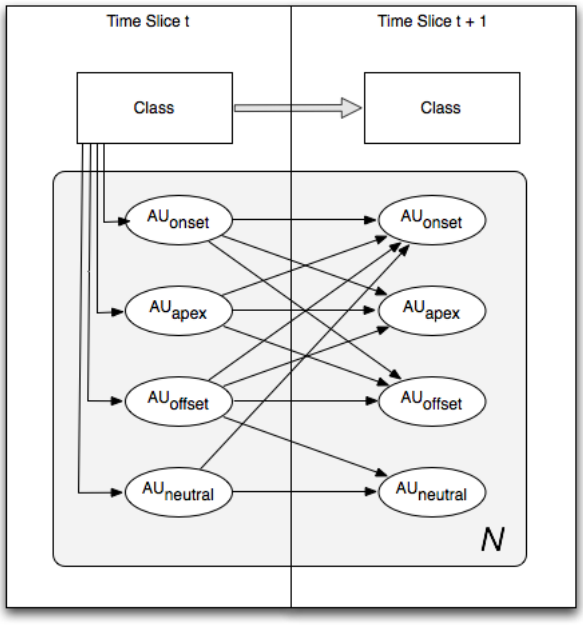

Figure 2: Dynamic Baysian Network for Temporal Phases

$$\mathbf{X}_t = \{Class, n * (AU_{onset}, AU_{apex}, AU_{offset}, AU_{neutral})\} \quad (17)$$

The first node represents the class of the video: whether or not the facial expression is posed or spontaneous. That is followed by N sets of four nodes, representing the four temporal phases of the AUs. The inter-slice relationships are defined by following the onset-apex-offset rules: only nodes that can logically follow each other are connected. The intra-slice relationships are defined by connecting the *Posed/Spontaneous* node to the temporal phases nodes, allowing the morphology to also be extrapolated within the network.

4.2 Apex Overlap and Symmetry

For Apex Overlap the DBN model consists of two boolean nodes per time slice: one node representing the class (posed or spontaneous), and one node to represent the apex overlap. For each time slice, we define the set $\mathbf{X}_t$:

$$\mathbf{X}_t = \{Class, ApexOverlap\} \quad (18)$$

An image of it can be seen in figure 3. For Symmetry the DBN model is exactly the same, only with the apex overlap node replaced with a symmetry node. For each time slice, we define the set $\mathbf{X}_t$:

$$\mathbf{X}_t = \{Class, Symmetry\} \quad (19)$$

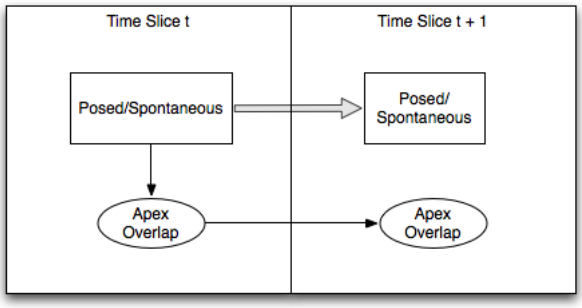

Figure 3: Dynamic Baysian Network for Apex Overlap

4.3 Total Duration and Speed

For Total Duration and Speed, the DBN model again consists of two boolean nodes per time slice: one node representing posed or spontaneous, and one node to represent the total duration/speed. For each time slice, we define the set $\mathbf{X}_t$:

$$\mathbf{X}_t = \{Class, TotalDuration\} \quad (20)$$

An image of it can be seen in figure 4. For Speed the DBN model is exactly the same, only with the total duration node replaced with a symmetry node. For each time slice, we define the set $\mathbf{X}_t$:

$$\mathbf{X}_t = \{Class, SpeedOnset\} \quad (21)$$

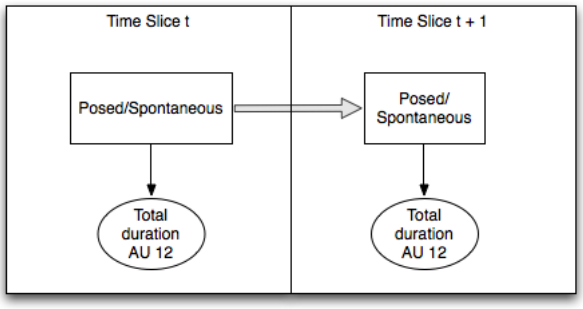

Figure 4: Dynamic Bayesian Network for Total Duration

5 The Dataset

To be able to evaluate the DBN models that we propose in this paper, we need a suitable set of data. The MMI Facial Expression Database is an online collection of video and audio recordings of subjects displaying facial expressions [16]. This continually growing database currently holds over 2900 videos and high-resolution still images of 75 subjects.

For our dataset we selected 50 videos from the MMI Facial Expression Database, half of which are of posed displays of happy and half of which are spontaneous displays of happy. The 25 posed displays are of 18 different subjects, who were all asked to express happiness. The 25 spontaneous displays are of 11 different subjects, who were each shown funny clips to induce happiness. Three subjects appear in both the posed and spontaneous sets. All videos were recorded with a frontal view, and under controlled lighting conditions. In figure 5 screen shots of two of the videos used can be seen.

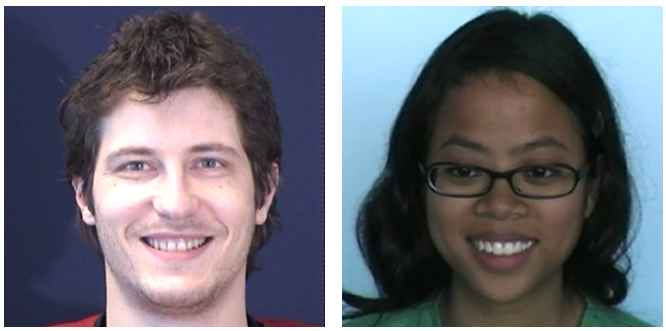

Figure 5: Screen shots of posed (left) and spontaneous (right) smiles from the MMI Facial Expression Database

The MMI Facial Expression Database provides the annotations of the event-coding for all videos in the dataset, but for our dataset we need the frame-by-frame level (onset-apex-offset) coding, indicating the temporal phase of an AU. For all missing oao-codings the videos were manually annotated using the ActionUnitCoding tool. Only AUs that occurred in more than 10% of the videos were considered (AU6, AU 7, AU 10 and AU 12).

In addition, we use the Patras-Pantic Particle Filtering with Factorized Likelihoods (PFFL) to track the necessary facial points [13]. In figure 6 the points can be seen, which are tracked with PFFL. We track two points on the mouth: R, the right mouth corner, and L, the left mouth corner. Additionally, we track the points on the inner eye corners and on the nose, to allow us to normalize the data. This is first done with intra-registration, removing all rigid head movements within the video, which is then followed by inter-registration, where the face is warped onto a predefined "normal" face, eliminating the inter-person variation of the face shape.

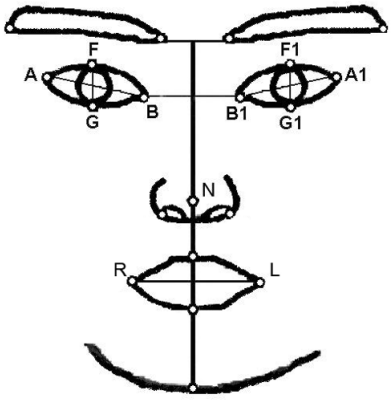

Figure 6: Facial points tracked with PFFL

6 Experiments

To evaluate the proposed method for distinguishing between posed and spontaneous facial expressions, we used the aforementioned dataset. For the Temporal Phases and Apex Overlap DBN models, all four available AUs were used (AU6, AU 7, AU 10 and AU 12). For the other facial features, only AU 12 is examined, but the internal threshold is varied to see which value performs best. The range of values of the thresholds β, γ and δ (eq. 12, 14 and 16) were chosen based off the density plots and ROC curves in figures 7 - 9.

Table 1 shows the average mean and standard deviation of the area under ROC curve performance, using five times 10-fold cross validation for the DBN models. For each video, a classification is made by first having the DBN calculate per frame the probability whether or not it is posed or spontaneous, and then averaging over the entire video for a final classification. The best performance for each facial feature is highlighted in bold. The DBN trained on the temporal phases of AU 6 and AU 12 combined performs the best classification. This confirms the psychology research: the zygomatic major (AU12) and the orbicularis oculi (AU 6) occurring in the same video is a likely indicator of a spontaneous smile, while the absence of the orbicularis oculi is a

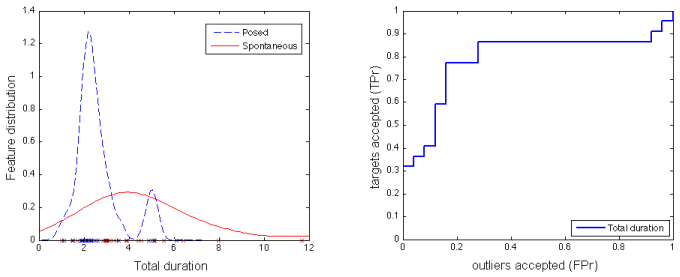

Figure 7: Density plot and ROC curve of the total duration for Posed and Spontaneous videos

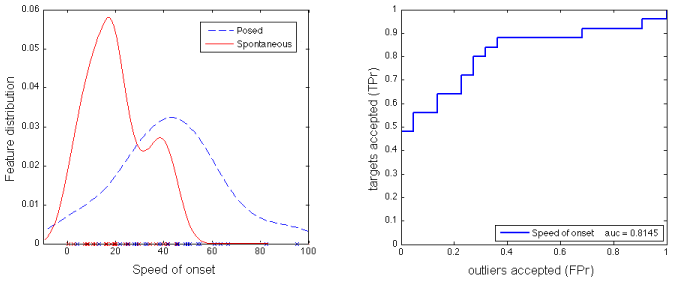

Figure 8: Density plot and ROC curve of the speed of onset for Posed and Spontaneous videos

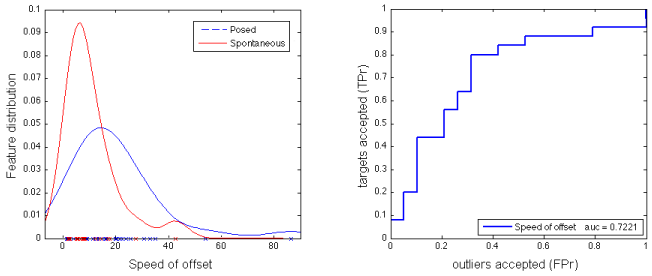

Figure 9: Density plot and ROC curve of the speed of offset for Posed and Spontaneous videos

strong indicator of a posed smile.

Overall the proposed temporal characteristics perform as desired, most being able to distinguish between posed and spontaneous facial expressions with fairly good accuracy. This confirms that the temporal dynamics of facial actions are important for the classification of posed and spontaneous emotions.

Finally, to investigate the combined effect of the DBN models, we merge the results of various combinations of DBNS, averaging the DBN results for each video. Table 2 shows the classification results using five times 10-fold cross validation for these combined classification. The combined

DBN Model	**AUC**
T1: Temporal Phases AU 6	0.80 (0.02)
T2: Temporal Phases AU 7	0.36 (0.04)
T3: Temporal Phases AU 10	0.50 (0.03)
T4: Temporal Phases AU 12	0.77 (0.02)
T5: Temporal Phases AU 6 & AU 10	0.74 (0.02)
T6: Temporal Phases AU 6 & AU 12	**0.85 (0.02)**
T7: Temporal Phases AU 10 & AU 12	0.73 (0.02)
T8: Temporal Phases AU 6 & AU 7	0.72 (0.01)
T9: Temporal Phases AU 7 & AU 10	0.55 (0.01)
T10: Temporal Phases AU 7 & AU 12	0.70 (0.02)
A1: Apex Overlap AU 6 & AU 10	**0.57 (0.03)**
A2: Apex Overlap AU 6 & AU 12	0.54 (0.02)
A3: Apex Overlap AU 10 & AU 12	0.51 (0.01)
A4: Apex Overlap AU 6 & AU 7	0.34 (0.04)
A5: Apex Overlap AU 7 & AU 10	0.34 (0.03)
A6: Apex Overlap AU 7 & AU 12	0.37 (0.03)
S1: Symmetry AU 12, $\alpha = 0.5$	0.55 (0.04)
S2: Symmetry AU 12, $\alpha = 0.6$	0.62 (0.03)
S3: Symmetry AU 12, $\alpha = 0.7$	**0.65 (0.02)**
S4: Symmetry AU 12, $\alpha = 0.8$	0.58 (0.03)
S5: Symmetry AU 12, $\alpha = 0.9$	0.62 (0.01)
D1: Total Duration AU 12, $\beta = 2.6$	0.76 (0.02)
D2: Total Duration AU 12, $\beta = 2.8$	0.78 (0.01)
D3: Total Duration AU 12, $\beta = 3.0$	**0.79 (0.01)**
D4: Total Duration AU 12, $\beta = 3.2$	0.75 (0.01)
D5: Total Duration AU 12, $\beta = 3.4$	0.72 (0.04)
O1: Speed of Onset AU 12, $\gamma = 25$	0.69 (0.01)
O2: Speed of Onset AU 12, $\gamma = 26$	**0.70 (0.01)**
O3: Speed of Onset AU 12, $\gamma = 27$	**0.70 (0.01)**
O4: Speed of Onset AU 12, $\gamma = 28$	0.69 (0.01)
O5: Speed of Onset AU 12, $\gamma = 29$	0.67 (0.01)
F1: Speed of Offset AU 12, $\delta = 8$	0.68 (0.02)
F2: Speed of Offset AU 12, $\delta = 9$	**0.72 (0.02)**
F3: Speed of Offset AU 12, $\delta = 10$	0.70 (0.02)
F4: Speed of Offset AU 12, $\delta = 11$	0.70 (0.02)
F5: Speed of Offset AU 12, $\delta = 12$	0.69 (0.02)

Table 1: Average AUC performances of the DBNs, using five times 10-fold cross validation, with the standard deviation over the five runs shown between brackets.

DBN C1 using all facial features achieves a much higher classification rate than the DBNs alone. This not only reconfirms that the temporal dynamics are essential to distinguishing between posed and spontaneous facial expressions, but that together they provide a more robust and precise classification. In classifiers C2 - C7 we examine what happens when one of the temporal characteristics is left out. The combined DBNs without Speed of Offset (C2) and without Symmetry (C5) performs just as well, indicating that these two features provide little extra contribution. Finally, in C8 we examine the combined DBN without both Speed of Offset and Symmetry and again attain the same performance rate as with them.

DBN Model	**AUC**
$T567 : T5 + T6 + T7$	0.94 (0.01)
$A123 : A1 + A2 + A3$	0.58 (0.02)
$C1 : T567 + A123 + S3 + D3 + O2 + F2$	**0.97 (0.00)**
$C2 : T567 + A123 + S3 + D3 + O2$	**0.97 (0.00)**
$C3 : T567 + A123 + S3 + D3 + F2$	0.97 (0.01)
$C4 : T567 + A123 + S3 + O2 + F2$	0.96 (0.01)
$C5 : T567 + A123 + D3 + O2 + F2$	**0.97 (0.00)**
$C6 : T567 + S3 + D3 + O2 + F2$	0.96 (0.00)
$C7 : A123 + S3 + D3 + O2 + F2$	0.95 (0.00)
$C8 : T567 + A123 + D3 + O2$	**0.97 (0.00)**

Table 2: Average AUC performances of the DBNs, using five times 10-fold cross validation.

7 Conclusions

In this paper we proposed using Dynamic Bayesian networks for distinguishing between posed and spontaneous facial expressions. Following the research in psychology, we built our system based on characteristics of the temporal dynamics in the face, and defined several facial features: morphology, apex overlap, symmetry, total duration, speed of onset and speed of offset. Based on these characteristics we built several DBN models to classify posed and spontaneous facial expressions. We attained a 97% performance rate when testing the system on 50 videos taken from the MMI database. The results confirmed research findings in psychology that temporal dynamics are essential for the classification of posed and spontaneous facial expressions.

From our study, it is made clear that the facial features of morphology, apex overlap, total duration and speed of onset are important for distinguishing between posed and spontaneous. Symmetry and the speed of offset, however, although both reported to be good indicators for posed and spontaneous facial recognition, did not appear to contribute to the DBN classification.

The DBN framework provides a suitable framework to represent the temporal dynamics of facial actions, allowing us to specify the relationships between and within the facial expression features. However, certain design choices for the DBNs were made in this study, and more research is needed to better understand what the most effective DBN model is. Firstly, the DBN models used only boolean, discrete nodes. Although they produced an adequate classification, it would be interesting to see how the DBNs perform with continuous nodes, representing certain facial features with probability density functions. Secondly, nodes were only connected from time slice t to time slice $t-1$. Nodes could take any number of nodes from past time slices into account, providing an internal memory to each time slice. Thirdly, when combining the facial features, we only examined high-level fusion, combining the DBNs based on those facial features. Further investigation and experimentation into how to fuse those facial features at a lower-level within a single DBN is strongly recommended. Modelling the facial features within a single DBN would allow relationships between features and establish dependencies that our model inherently disregarded.

In summary, DBNs are capable of distinguishing between posed and spontaneous facial expressions, and the temporal dynamics are shown to be key in the classification. More extensive experiments are needed though to fully investigate what the best DBN model is. Future research should also examine the effect of a larger dataset and automatically labelled temporal phases data. Finally it needs to be researched whether other facial expressions beyond smiles also benefit from the same temporal dynamics for posed versus spontaneous facial expression recognition.

References

[1] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *IEEE Interantional Conference on Automatic Face and Gesture Recognition*, 2006.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[3] J. Cohn, L. I. Reed, Z. Ambadar, J. Xiao, and T. Moriyama. Automatic analysis and recognition of brow actions spontaneous facial behavior. *Proc. IEEE Int'l Conf. Systems, Man & Cybernetics*, pages 610–616, 2004.

[4] J. F. Cohn and K. L. Schmidt. The timing of facial motion in posed and spontaneous smiles. *J. Wavelets, Multi-resolution & Information Processing*, 2004.

[5] P. Ekman. Darwin, deception and facial expression. *Annals New York Academy of sciences*, pages 205–221, 2003.

[6] P. Ekman and E. Eds. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System*. Oxford University Press, 2005.

[7] P. Ekman and W. V. Friesen. Felt, false and miserable smiles. In *Journal of Nonverbal Behavior 6(4)*, pages 238–252, 1982.

[8] P. Ekman, J. C. Hager, and W. V. Friesen. The symmetry of emotional and deliberate facial actions. In *Psychophysiology, 18*, pages 101–106, March 1981.

[9] U. Hess and R. E. Kleck. Differentiating emotion elicited and deliberate emotional facial expression. In *European Journal of Social Psychology*, volume Vol. 20, pages 369–385, 1990.

[10] K. P. Murphy. The bayes net toolbox for matlab. *Computing Science and Statistics*, 2001.

[11] K. P. Murphy. Dynamic bayesian networks: Representation, inference and learning. Master's thesis, University of California, 2002.

[12] M. Pantic and L. J. Rothkrantz. Towards an affect-sensitive multimodal human-computer interaction. In *Proceedings of the IEEE, Vol. 91, No. 9*, 2003.

[13] I. Patras and M. Pantic. Particle filtering with factorized likelihoods for tracking facial features. In *Proceedings of IEEE Int'l Conf. Face and Gesture Recognition (FG'04)*, 2004.

[14] S. J. Russell, P. Norvig, J. F. Candy, J. M. Malik, and D. D. Edwards. *Artificial Intelligence: a modern approach*. Prentice-Hall, Inc., 1996.

[15] M. Valstar, H. Gunes, and M. Pantic. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces*, 2007.

[16] M. Valstar and M. Pantic. Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In *Proceedings of Int'l Conf. Language Resources and Evaluation, Workshop on EMOTION*, 2010.