Are You Really Smiling at Me? Spontaneous versus Posed Enjoyment Smiles

Hamdi Dibeklioğlu[‡], Albert Ali Salah[§], and Theo Gevers[‡]

[‡]Intelligent Systems Lab Amsterdam, University of Amsterdam, Amsterdam, The Netherlands
{h.dibeklioglu,th.gevers}@uva.nl
[§]Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey
salah@boun.edu.tr

Abstract. Smiling is an indispensable element of nonverbal social interaction. Besides, automatic distinction between spontaneous and posed expressions is important for visual analysis of social signals. Therefore, in this paper, we propose a method to distinguish between spontaneous and posed enjoyment smiles by using the dynamics of eyelid, cheek, and lip corner movements. The discriminative power of these movements, and the effect of different fusion levels are investigated on multiple databases. Our results improve the state-of-the-art. We also introduce the largest spontaneous/posed enjoyment smile database collected to date, and report new empirical and conceptual findings on smile dynamics. The collected database consists of 1240 samples of 400 subjects. Moreover, it has the unique property of having an age range from 8 to 76 years. Large scale experiments on the new database indicate that eyelid dynamics are highly relevant for smile classification, and there are age-related differences in smile dynamics.

Key words: Face analysis, smile classification, affective computing

1 Introduction

Human facial expressions are indispensable elements of non-verbal communication. Since faces can reveal the mood or the emotional feeling of a person, automatic understanding and interpretation of facial expressions provide a natural way to interact with computers. In recent studies, analysis of spontaneous facial expressions have gained more interest. For social interaction analysis, it is necessary to distinguish spontaneous (felt) expressions from the posed (deliberate) ones. Spontaneous expressions can reveal states of attention, agreement and interest, as well as deceit. The foremost facial expression for spontaneity analysis is the smile, as it is the most frequently performed expression, and used for signaling enjoyment, embarrassment, politeness, etc. [1]. It is also used to mask other emotional expressions, since it is the easiest emotional facial expression to pose voluntarily [2].

Several characteristics of spontaneous and posed smiles, such as symmetry, speed, and timing are analyzed in the literature [3], [4], [5]. Their findings suggest that different facial regions contribute differently to the classification of smiles. In this paper, we assess the facial dynamics for different face regions, and demonstrate that the eye region contains the most useful information for this problem. Our method combines

region-specific smile dynamics (duration, amplitude, speed, acceleration, etc.) with eyelid movements to detect the genuineness of enjoyment smiles.

Our contributions are: 1) A region-specific analysis of facial feature movements under various conditions; 2) An accurate smile classification method which outperforms the state-of-the-art methods; 3) New empirical findings on age related differences in smile expression dynamics; 4) The largest spontaneous/posed enjoyment smile database in the literature for detailed and precise analysis of enjoyment smiles. The database, its evaluation protocols and annotations are made available to the research community.

2 Related Work

The smile is the easiest emotional facial expression to pose voluntarily [2]. Broadly, a smile can be identified as the upward movement of the mouth corners, which corresponds to Action Unit 12 (AU12) in the facial action coding system (FACS) [6]. In terms of anatomy, the *zygomatic major* muscle contracts and raises the corners of the lips during a smile [4]. In terms of dynamics, smiles are composed of three non-overlapping phases; the onset (neutral to expressive), apex, and offset (expressive to neutral), respectively. Ekman individually identified 18 different smiles (such as enjoyment, fear, miserable, embarrassment, listener response smiles) by describing the specific visual differences on the face and indicating the accompanying action units [2].

Smiles of joy are called Duchenne smiles (D-smiles) in honor of Guillaume Duchenne, who did early experimental work on smiles. A good indicator for the D-smile is the contraction of the *orbicularis oculi, pars lateralis* muscle that raises the cheek, narrows the eye aperture, and forms wrinkles on the external side of the eyes. This activation corresponds to Action Unit 6 and is called the Duchenne marker (D-marker) in the literature. However, new empirical findings questioned the reliability of the D-marker [7]. Recently, it has been shown that *orbicularis oculi, pars lateralis* can be active or inactive under both spontaneous and posed conditions with similar frequencies [8]. On the other hand, untrained people consistently use the D-marker to recognize spontaneous and posed enjoyment smiles [9].

Symmetry is also potentially informative to distinguish spontaneous and posed enjoyment smiles [4]. In [3], it is claimed that spontaneous enjoyment smiles are more symmetrical than posed ones. Later studies reported no significant difference [7]. This study has also failed to find significant effects of symmetry.

In the last decade, dynamical properties of smiles (such as duration, speed, and amplitude of smiles; movements of head and eyes) received attention as opposed to morphological cues to discriminate between spontaneous and posed smiles. In [10], Cohn *et al.* analyze correlations between lip-corner displacements, head rotations, and eye motion during spontaneous smiles. In another study, Cohn and Schmidt report that spontaneous smiles have smaller onset amplitude of lip corner movement, but a more stable relation between amplitude and duration [5]. Furthermore, the maximum speed of the smile onset is higher in posed samples and posed eyebrow raises have higher maximum speed and larger amplitude, but shorter duration than spontaneous ones [7].

In [5], Cohn and Schmidt propose a system which distinguishes spontaneous and deliberate enjoyment smiles by a linear discriminant classifier using *duration*, *amplitude*,

and $\frac{duration}{amplitute}$ measures of smile onsets. They analyze the significance of the proposed features and show that the amplitude of the lip corner movement is a strong linear function of duration in spontaneous smile, but not in deliberate ones. In [11], Valstar *et al.* propose a multimodal system to classify posed and spontaneous smiles. GentleSVM-Sigmoid classifier is used with the fusion of shoulder, head and inner facial movements.

In [12], Pfister *et al.* propose a spatiotemporal method using both natural and infrared face videos to discriminate between spontaneous and posed facial expressions. By enabling the temporal space and using the image sequence as a volume, they extend the Completed Local Binary Patterns (CLBP) texture descriptor into the spatio-temporal CLBP-TOP features for this task.

Recently, Dibeklioğlu *et al.* have proposed a system which uses eyelid movements to classify spontaneous and posed enjoyment smiles [13], where distance-based and angular features are defined in terms of changes in eye aperture. Several classifiers are compared and the reliability of eyelid movements are shown to be superior to that of the eyebrows, cheek, and lip movements for smile classification.

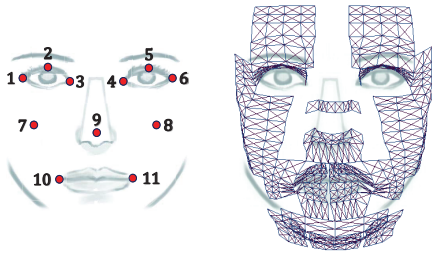
In conclusion, the most relevant facial cues for smile classification in the literature are 1) D-marker, 2) the symmetry, and 3) the dynamics of smiles. Instead of analyzing these facial cues separately, in this paper, the aim is to use a more generic descriptor set which can be applied to different facial regions to enhance the indicated facial cues with detailed dynamic features. Additionally, we focus on the dynamical characteristics of eyelid movements (such as duration, amplitude, speed, and acceleration), instead of simple displacement analysis, motivated by the findings of [5] and [13].

3 Method

In this section, details of the proposed spontaneous/posed enjoyment smile classification system will be summarized. The flow of the system is as follows. Facial fiducial points are located in the first frame, and tracked during the rest of the smile video. These points are used to calculate displacement signals of eyelids, cheeks, and lip corners. Onset, apex, and offset phases of the smile are estimated using the mean displacement of the lip corners. Afterwards, descriptive features for eyelid, cheek, and lip corner movements are extracted from each phase. After a feature selection procedure, the most informative features with minimum dependency are used to train the Support Vector Machine (SVM) classifiers.

3.1 Facial Feature Tracking

To analyze the facial dynamics, 11 facial feature points (eye corners, center of upper eyelids, cheek centers, nose tip, lip corners) are tracked in the videos (see Fig. 1(a)). Each point is initialized in the first frame of the videos for precise tracking and analysis. In our system, we use the piecewise Bézier volume deformation (PBVD) tracker, which is proposed by Tao and Huang [14] (see Fig. 1(b)). While this method is relatively old, we have introduced improved methods for its initialization, and it is fast and robust with accurate initialization. The generic face model consists of 16 surface patches. To form a continuous and smooth model, these patches are embedded in Bézier volumes.



(a) (b)

Fig. 1. (a) Used facial feature points with their indices and (b) the 3D mesh model

3.2 Feature Extraction

Three different face regions (eyes, cheeks, and mouth) are used to extract descriptive features. First of all, tracked 3D coordinates of the facial feature points ℓ_i (see Fig. 1(a)) are used to align the faces in each frame. We estimate the 3D pose of the face, and normalize the face with respect to roll, yaw, and pitch rotations. Since three non-colinear points are enough to construct a plane, we use three stable landmarks (eye centers and nose tip) to define a plane $\mathcal{P}$. Eye centers are defined as middle points between inner and outer eye corners as $c_1 = \frac{\ell_1+\ell_3}{2}$ and $c_2 = \frac{\ell_4+\ell_6}{2}$. Angles between the positive normal vector $\mathcal{N}_\mathcal{P}$ of $\mathcal{P}$ and unit vectors U on X (horizontal), Y (vertical), and Z (perpendicular) axes give the relative head pose as follows:

$$\theta = \arccos \frac{U.\mathcal{N}_\mathcal{P}}{\|U\| \, \|\mathcal{N}_\mathcal{P}\|}, \text{ where } \mathcal{N} = \overrightarrow{\ell_9 c_2} \times \overrightarrow{\ell_9 c_1}. \tag{1}$$

In Equation 1, $\overrightarrow{\ell_9 c_2}$ and $\overrightarrow{\ell_9 c_1}$ denote the vectors from point ℓ_9 to points c_2 and c_1, respectively. $\|U\|$ and $\|\mathcal{N}_\mathcal{P}\|$ are the magnitudes of U and $\mathcal{N}_\mathcal{P}$ vectors. According to the face geometry, Equation 1 can estimate the exact roll (θ_z) and yaw (θ_y) angles of the face with respect to the camera. If we assume that the face is approximately frontal in the first frame, then the actual pitch angles (θ'_x) can be calculated by subtracting the initial value. Once the pose of the head is estimated, tracked points are normalized with respect to rotation, scale, and translation as follows:

$$\ell'_i = \left[\ell_i - \frac{c_1 + c_2}{2} \right] R_x(-\theta'_x) R_y(-\theta_y) R_z(-\theta_z) \frac{100}{\rho(c1, c2)}, \tag{2}$$

where ℓ'_i is the aligned point and R_x, R_y, and R_z denote the 3D rotation matrices for the given angles. $\rho()$ denotes the Euclidean distance. On the normalized face, the middle point between eye centers is located at the origin and the interocular distance (distance between eye centers) is set to 100 pixels. Since the normalized face is approximately frontal with respect to the camera, we ignore the depth (Z) values of the normalized feature points ℓ'_i, and denote them as l_i.

After the normalization, onset, apex, and offset phases of the smile can be detected using the approach proposed by Schmidt *et al.* [15], by calculating the amplitude of the

smile as the distance of the right lip corner to the lip center during the smile. Differently from [15], we estimate the smile amplitude as the mean amplitude of right and left lip corners, normalized by the length of the lip. Let $\mathcal{D}_{\text{lip}}(t)$ be the value of the mean amplitude signal of the lip corners in the frame t. It is estimated as:

$$\mathcal{D}_{\text{lip}}(t) = \frac{\rho(\frac{l^1_{10}+l^1_{11}}{2}, l^t_{10}) + \rho(\frac{l^1_{10}+l^1_{11}}{2}, l^t_{11})}{2\rho(l^1_{10}, l^1_{11})}, \tag{3}$$

where l^t_i denotes the 2D location of the i^{th} point in frame t. The longest continuous increase in $\mathcal{D}_{\text{lip}}$ is defined as the onset phase. Similarly, the offset phase is detected as the longest continuous decrease in $\mathcal{D}_{\text{lip}}$. The phase between the last frame of the onset and the first frame of the offset defines the apex.

To extract features from the eyelids and the cheeks, additional amplitude signals are computed. We estimate the (normalized) eyelid aperture $\mathcal{D}_{\text{eyelid}}$ and cheek displacement $\mathcal{D}_{\text{cheek}}$ as follows:

$$\mathcal{D}_{\text{eyelid}}(t) = \frac{\kappa(\frac{l^t_1+l^t_3}{2}, l^t_2)\rho(\frac{l^t_1+l^t_3}{2}, l^t_2) + \kappa(\frac{l^t_4+l^t_6}{2}, l^t_5)\rho(\frac{l^t_4+l^t_6}{2}, l^t_5)}{2\rho(l^t_1, l^t_3)}, \tag{4}$$

$$\mathcal{D}_{\text{cheek}}(t) = \frac{\rho(\frac{l^1_7+l^1_8}{2}, l^t_7) + \rho(\frac{l^1_7+l^1_8}{2}, l^t_8)}{2\rho(l^1_7, l^1_8)}, \tag{5}$$

where $\kappa(l_i, l_j)$ denotes the relative vertical location function, which equals to -1 if l_j is located (vertically) below l_i on the face, and 1 otherwise. $\mathcal{D}_{\text{lip}}$, $\mathcal{D}_{\text{eyelid}}$, and $\mathcal{D}_{\text{cheek}}$ are hereafter referred to as amplitude signals. In addition to the amplitudes, speed $\mathcal{V}$ and acceleration $\mathcal{A}$ signals are extracted by computing the first and the second derivatives of the amplitudes, respectively.

In summary, description of the used features and the related facial cues with those are given in Table 1. Note that the defined features are extracted separately from each phase of the smile. As a result, we obtain three feature sets for each of the eye, mouth and cheek regions. Each phase is further divided into increasing ($^+$) and decreasing ($^-$) segments, for each feature set. This allows a more detailed analysis of the feature dynamics.

In Table 1, signals symbolized with superindex ($^+$) and ($^-$) denote the segments of the related signal with continuous increase and continuous decrease, respectively. For example, $\mathcal{D}^+$ pools the increasing segments in $\mathcal{D}$. η defines the length (number of frames) of a given signal, and ω is the frame rate of the video. $\mathcal{D}_L$ and $\mathcal{D}_R$ define the amplitudes for the left and right sides of the face, respectively. For each face region, three 25-dimensional feature vectors are generated by concatenating these features.

In some cases, features cannot be calculated. For example, if we extract features from the amplitude signal of the lip corners $\mathcal{D}_{\text{lip}}$ using the onset phase, then decreasing segments will be an empty set ($\eta(\mathcal{D}^-) = 0$). For such exceptions, all the features describing the related segments are set to zero.

3.3 Feature Selection and Classification

To classify spontaneous and posed smiles, individual SVM classifiers are trained for different face regions. As described in Section 3.2, we extract three 25-dimensional

Table 1. Definitions of the extracted features, and the related facial cues with those. The relation with D-marker is only valid for eyelid features

Feature	Definition	Related Cue(s)
Duration:	$\left[\ \frac{\eta(\mathcal{D}^+)}{\omega}\ ,\ \frac{\eta(\mathcal{D}^-)}{\omega}\ ,\ \frac{\eta(\mathcal{D})}{\omega}\ \right]$	Dynamics
Duration Ratio:	$\left[\ \frac{\eta(\mathcal{D}^+)}{\eta(\mathcal{D})}\ ,\ \frac{\eta(\mathcal{D}^-)}{\eta(\mathcal{D})}\ \right]$	Dynamics
Maximum Amplitude:	$\max(\mathcal{D})$	Dynamics, D-marker
Mean Amplitude:	$\left[\ \frac{\sum \mathcal{D}}{\eta(\mathcal{D})}\ ,\ \frac{\sum \mathcal{D}^+}{\eta(\mathcal{D}^+)}\ ,\ \frac{\sum \lvert\mathcal{D}^-\rvert}{\eta(\mathcal{D}^-)}\ \right]$	Dynamics, D-marker
STD of Amplitude:	$\mathrm{std}(\mathcal{D})$	Dynamics
Total Amplitude:	$\left[\ \sum \mathcal{D}^+\ ,\ \sum \lvert\mathcal{D}^-\rvert\ \right]$	Dynamics
Net Amplitude:	$\sum \mathcal{D}^+ - \sum \lvert\mathcal{D}^-\rvert$	Dynamics
Amplitude Ratio:	$\left[\ \frac{\sum \mathcal{D}^+}{\sum \mathcal{D}^+ + \sum \lvert\mathcal{D}^-\rvert}\ ,\ \frac{\sum \lvert\mathcal{D}^-\rvert}{\sum \mathcal{D}^+ + \sum \lvert\mathcal{D}^-\rvert}\ \right]$	Dynamics
Maximum Speed:	$\left[\ \max(\mathcal{V}^+)\ ,\ \max(\lvert\mathcal{V}^-\rvert)\ \right]$	Dynamics
Mean Speed:	$\left[\ \frac{\sum \mathcal{V}^+}{\eta(\mathcal{V}^+)}\ ,\ \frac{\sum \lvert\mathcal{V}^-\rvert}{\eta(\mathcal{V}^-)}\ \right]$	Dynamics
Maximum Acceleration:	$\left[\ \max(\mathcal{A}^+)\ ,\ \max(\lvert\mathcal{A}^-\rvert)\ \right]$	Dynamics
Mean Acceleration:	$\left[\ \frac{\sum \mathcal{A}^+}{\eta(\mathcal{A}^+)}\ ,\ \frac{\sum \lvert\mathcal{A}^-\rvert}{\eta(\mathcal{A}^-)}\ \right]$	Dynamics
Net Ampl., Duration Ratio:	$\frac{\left(\sum \mathcal{D}^+ - \sum \lvert\mathcal{D}^-\rvert\right)\omega}{\eta(\mathcal{D})}$	Dynamics
Left/Right Ampl. Difference:	$\frac{\lvert\sum \mathcal{D}_L - \sum \mathcal{D}_R\rvert}{\eta(\mathcal{D})}$	Symmetry

feature vectors for each face region. To deal with feature redundancy, we use Min-Redundancy Max-Relevance (mRMR) algorithm to select discriminative features [16]. mRMR is an incremental method for minimizing the redundancy while selecting the most relevant information as follows:

$$\max_{f_j \in F - S_{m-1}} \left[I\left(f_j, c\right) - \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I\left(f_j, f_i\right) \right], \tag{6}$$

where I shows the mutual information function and c indicates the target class. F and S_{m-1} denote the feature set, and the set of $m-1$ features, respectively.

During the training of our system, both individual feature vectors for onset, apex, offset phases, and a vector with their fusion are generated. The most discriminative features on each of the generated feature sets are selected using mRMR. Minimum classification error on a separate validation set is used to determine the most informative facial region, and the most discriminative features on the selected region. Similarly, to optimize the SVM configuration, different kernels (linear, polynomial, and radial basis function (RBF)) with different parameters (size of RBF kernel, degree of poly-

nomial kernel) are tested on the validation set and the configuration with the minimum validation error is selected. The test partition of the dataset is not used for parameter optimization.

4 Database

4.1 UvA-NEMO Smile Database

We have recently collected the UvA-NEMO Smile Database[1] to analyze the dynamics of spontaneous/posed enjoyment smiles. This database is composed of videos (in RGB color) recorded with a Panasonic HDC-HS700 3MOS camcorder, placed on a monitor, at approximately 1.5 meters away from the recorded subjects. Videos were recorded with a resolution of 1920×1080 pixels at a rate of 50 frames per second under controlled illumination conditions. Additionally, a color chart is present on the background of the videos for further illumination and color normalization (See Fig. 2).

The database has 1240 smile videos (597 spontaneous, 643 posed) from 400 subjects (185 female, 215 male), making it the largest smile database in the literature so far. Ages of subjects vary from 8 to 76 years. 149 subjects are younger than 18 years (235 spontaneous, 240 posed). 43 subjects do not have spontaneous smiles and 32 subjects have no posed smile samples. (See Fig. 3 for age and gender distributions).

Fig. 2. Spontaneous (top) and posed (bottom) enjoyment smiles from the UvA-NEMO Smile Database

For posed smiles, each subject was asked to pose an enjoyment smile as realistically as possible, sometimes after being shown the proper way in a sample video. Short, funny video segments were used to elicit spontaneous enjoyment smiles. Approximately five minutes of recordings were made per subject, and genuine smiles were segmented. For

[1] This research was part of Science Live, the innovative research programme of science center NEMO that enables scientists to carry out real, publishable, peer-reviewed research using NEMO visitors as volunteers. See http://www.uva-nemo.org on how to obtain the UvA-NEMO Smile Database.

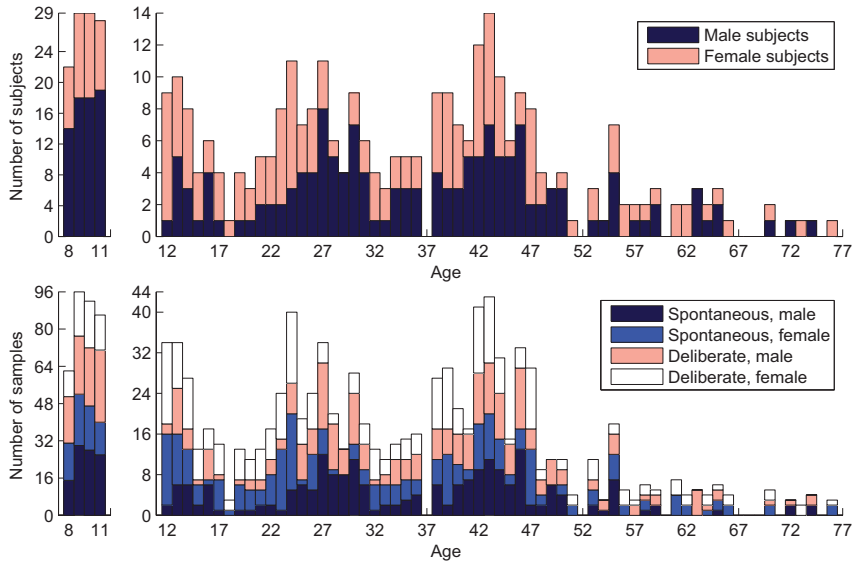


Fig. 3. Age and gender distributions for the subjects (top), and for the smiles (bottom) in the UvA-NEMO Smile Database

each subject, a balanced number of spontaneous and posed smiles were selected and annotated by seeking consensus of two trained annotators. Segments start/end with neutral or near-neutral expressions. The mean duration of the smile samples is 3.9 seconds ($\sigma = 1.8$). Average interocular distance on the database is approximately 200 pixels (estimated by using the tracked landmarks). 50 subjects wear eyeglasses.

4.2 Existing Smile Databases

Facial expression databases in the literature rarely contain spontaneous smiles. We have used several existing databases to report results with the proposed method. Table 2 shows a comparative overview of publicly available smile databases. *BBC Smile Dataset*[2] was gathered from "Spot the fake smile" test on the BBC website, with 10 spontaneous and 10 posed smile videos, each from a different subject, and each starting and ending with a neutral face. *MMI Facial Expression Database* [17] is not specifically gathered for smile classification, but includes 74 posed smiles from 30 subjects, as well as spontaneous smiles from 25 subjects. *SPOS Corpus* [12] contains natural color and infrared videos of 66 spontaneous and 14 posed smiles from seven subjects. At the moment only the onsets of expressions are available, but the entire database will be opened to the public. *USTC-NVIE Database* [18] consists of images and videos of six basic facial expressions and neutral faces. Images are recorded in both natural color and infrared, simultaneously, under three different illumination conditions. NVIE database has two parts; spontaneous part includes image sequences of onset phases, where posed

[2] http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/

Table 2. Overview of databases with spontaneous/posed smile content. Note that the posed part of USTC-NVIE includes only the most expressive (apex) images

Database	Participants	Resolution & Frame Rate	Age Annotation (range)
BBC	20	314×286 pixels @25Hz	No (unknown)
MMI [17]	55	720×576 pixels @25Hz 576×720 pixels @25Hz 640×480 pixels @29Hz	Yes (19–64 years)
SPOS [12]	7	640×480 pixels @25Hz	No (unknown)
USTC-NVIE [18]	148	640×480 pixels @30Hz 704×480 pixels @30Hz	No (17–31 years)
UvA-NEMO	400	1920×1080 pixels @50Hz	Yes (8–76 years)

part consists of only the most expressive (apex) images. The database has 302 spontaneous smiles (image sequences of onset phases) from 133 subjects, and 624 posed smiles (single apex images) from 104 subjects.

5 Experimental Results

We use the UvA-NEMO smile database to evaluate our system, but report further results on BBC and SPOS corpora in Section 5.4. We use a two level 10-fold cross-validation scheme: Each time a test fold is separated, a 9-fold cross-validation is used to train the system, and parameters are optimized without using the test partition. There is no subject overlap between folds. For classification, linear SVM is found to perform better than polynomial and RBF alternatives. For detailed analysis of the features and the system, tracking is initialized by manually annotated facial landmarks. Additionally, results with automatic initialization are also given in Section 5.4.

5.1 Assessment of Facial Regions and Feature Selection

To evaluate the discriminative power of eye, cheek, and mouth regions for smile classification, we use features of onset, apex, and offset phases of all regions, individually. Additionally, features of all phases are concatenated and tested for each region (shown as *All* in Fig. 4). Feature selection reduces the number of feature dimensions and increases the correct classification rates for each feature set, except for apex features of eyelid and lip corners, and offset features of cheeks (Fig. 4). Decrease in the accuracy for these three feature sets are around 1% (absolute), where the feature selection increases the accuracy with approximately 3% (absolute) on average. Since these results confirm the usefulness of feature selection, it is used in the remainder of this section.

When we analyze the results with feature selection, it is seen that the best accuracy, for individual phases, is achieved by the onset features of lip corners (82.58%). However, the discriminative power of apex and offset phases of lip corners do not reach those of the eyelids and cheeks. The most reliable apex and offset features are obtained

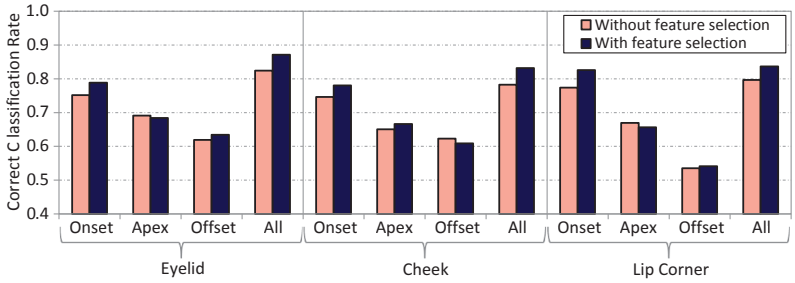


Fig. 4. Effect of feature selection on correct classification rates for different facial regions

from the eyelids, which provide the highest correct classification rate (87.10%) when the features of all phases are concatenated. Lip corners (83.63%) and cheeks (83.15%) follow the eyelids. Combined eyelid features have the minimum validation error in this experiment.

5.2 Assessment of Fusion Strategies

Three different fusion strategies (early, mid-level, and late) are defined and evaluated. Each fusion strategy enables feature selection before classification. In early fusion, features of onset, apex, and offset of all regions are fused into one low-abstraction vector and classified by a single classifier. Mid-level fusion concatenates features of all phases for each region, separately. Constructed feature vectors are individually classified by SVMs and the classifier outputs are fused (either by SUM or PRODUCT rule, or by voting[3]). In the late fusion scheme, feature sets of onset, apex, and offset for all facial regions are individually classified by SVMs and fused.

As shown in Fig. 5 (a), mid-level fusion provides the best performance (88.87% with voting), followed by early and late fusion, respectively. Elimination of redundant information by feature selection after low-level feature abstraction on each region, separately, and following higher level of abstraction for classification on different facial regions can explain the high performance of mid-level fusion.

5.3 Effect of Age

The features on which we base our analysis may depend on the age of the subjects. In this section, we split the UvA-NEMO smile database into two partitions as *young people* ($age < 18$ years), and *adults* ($age \geq 18$ years), and all training and evaluation is repeated separately for the two partitions of the database. Fig. 5 (b) shows the classification accuracies. Regional performances are given using the fused (onset, apex, offset) features of the related region. Mid-level fusion (voting) accuracies are also given in Fig. 5 (b).

[3] The SUM and PRODUCT rules fuse the computed posterior probabilities for the target classes of different classifiers. To estimate these posterior probabilities, sigmoids of SVM output distances are used.

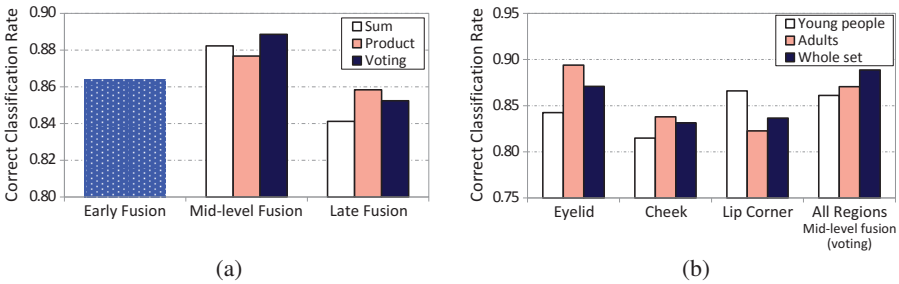


Fig. 5. (a) Effect of different fusion strategies and (b) age on correct classification rates

Our results show that eyelid features perform better on *adults* (as well as on the *whole set*) than cheeks and lip corners. For *adults*, eyelid features reach an accuracy of 89.41%, where features of cheeks and lip corners have an accuracy of 83.79% and 82.22%, respectively. However, the correct classification rate for *young people* with eyelid features is approximately 5% and 3% less than on the *adults* and *whole set*, respectively. Lip corner features provide the most reliable classification for *young people* with an accuracy of 86.53%, which also have the minimum validation error for this partition. Additionally, results show that fusion does not increase the performance for *adults* and *young people*, individually, and the highest correct classification rate is achieved on the *whole set*.

5.4 Comparison with Other Methods

We compare our method with the state-of-the-art smile classification systems proposed in the literature [5], [13], [12] by evaluating them on the UvA-NEMO database with the same experimental protocols, as well as on BBC and SPOS. Results for [12] are given by using only natural texture images. For a fair comparison, all methods are tested by using the piecewise Bézier volume tracker [14] and the tracker is initialized **automatically** by the method proposed in [19]. Correct classification rates are given in Table 3. Results show that the proposed method outperforms the state-of-the-art methods. Mid-level fusion with voting provides an accuracy of 87.02%, which is 9.76% (absolute) higher than the performance of the method proposed in [5]. Using only eyelid features decreases the correct classification rate by only 1.29% (absolute) in comparison to mid-level fusion. This confirms the reliability of eyelid movements and the discriminative power of the proposed dynamical eyelid features to distinguish between types of smiles.

Our system with eyelid features has an 85.73% accuracy, significantly higher than that of [13] (71.05%), which uses only eyelid movement features without any temporal segmentation. This shows the importance of temporal segmentation. The results of [12] with 73.06% classification rate is less than the accuracy of our method with only onset features, and shows that spatiotemporal features are not as reliable as facial dynamics.

Since [5] relies on solely the onset features of lip corners, we also tested our method with onset features of lip corners, and obtained an accuracy of 80.73% (compared to [5]'s 77.26%). We conclude that using automatically selected features from a large

Table 3. Correct classification rates on the BBC, SPOS, and UvA-NEMO databases

Method	Correct Classification Rate (%)		
	BBC	SPOS	UvA-NEMO
Proposed, Eyelid Features	85.00	72.50	85.73
Proposed, Mid-level fusion (voting)	**90.00**	**75.00**	**87.02**
Cohn & Schmidt [5]	75.00	72.50	77.26
Dibeklioğlu et al. [13]	85.00	66.25	71.05
Pfister et al. [12]	70.00	67.50	73.06

pool of informative features serves better than enabling a few carefully selected measures for this problem. Manually selected features may also show less generalization power across different (database-specific) recording conditions.

It is important to note that the proposed method uses solely onset features on SPOS corpus, since it has only onset phases of smiles. We have observed that spontaneous smiles are generally classified better than posed ones for all methods. One possible explanation is that dynamical facial features have more variance in posed smiles. Subsequently, class boundaries of spontaneous smiles are more defined, and this leads to a higher accuracy.

6 Discussion

In our experiments, onset features of lip corners perform best for individual phases. This result is consistent with the findings of Cohn *et al.* [5]. However, when onset, apex, and offset phases are fused, the eyelid movements are more descriptive than those of cheeks and lip corners for enjoyment smile classification.

On UvA-NEMO, the best fusion scheme increases the correct classification rate by only 1.29% (absolute) with respect to the accuracy of eyelid features. This finding supports our motivation and confirms the discriminative power and the reliability of eyelid movements to classify enjoyment smiles. However, it is important to note that temporal segmentation of the smiles are performed by using lip corner movements, which means that additional information from the movements of lip corners is leveraged.

Highly significant ($p < 0.001$) feature differences (of selected features) between *adults* and *young people* are obtained. For both spontaneous and posed smiles, maximum and mean apertures of eyes are larger for *adults*. During onset, both amplitude of eye closure and closure speed are higher for *young people*. During offset, amplitude and speed of eye opening are higher for *young people*. When we analyze the significance levels of the most selected features for smile classification, we see that the size of the eye aperture is smaller during spontaneous smiles. However, many subjects in UvA-NEMO lower their eyelids also during posed smiles. This result is consistent with the findings of Krumhuber and Manstead [8], which indicates that D-marker can exist during both spontaneous and posed enjoyment smiles.

Another important finding is that the speed and acceleration of eyelid movements are higher for posed smiles. As a result, since faster eyelid movements of *young people*

cause confusion with posed smiles, the classification accuracy with eyelid features is higher for *adults*. Similarly, features extracted from the cheek region perform better for *adults*, since cheek movements of *adults* are slower and more stationary. The duration of spontaneous smiles are longer than posed ones, but the lip corner movement for posed smiles is faster (also in terms of acceleration). This improves the accuracy of classification with lip corner features in favor of *young people*, since the lip corner movements of *young people* are significantly faster than *adults* during posed smiles.

Since eyelid and cheek features are reliable in *adults* as opposed to lip corners in *young people*, fusion on a specific age group decreases the accuracy compared to the performance on the *whole set*. Lastly, there is no significant symmetry difference (in terms of amplitude) between spontaneous and posed smiles (as in [20], [7]) or between *young people* and *adults*. More detailed analysis of age related smile dynamics is given in [21], which uses the proposed features (facial dynamics) for age estimation.

7 Conclusions

In this paper, we have presented a smile classifier that can distinguish posed and spontaneous enjoyment smiles with high accuracy. The method is based on the hypothesis that dynamics of eyelid movements are reliable cues for identifying posed and spontaneous smiles. Since the movements of eyelids are complex to analyze (because of blinks and continuous change in eye aperture), novel features have been proposed to describe dynamics of eyelid movements in detail. The proposed features also incorporate facial cues previously used in the literature (D-marker, symmetry, and dynamics) for classification of smiles, and can be generalized to any facial region.

We have introduced the largest spontaneous/posed enjoyment smile database (1240 samples of 400 subjects, ages of subjects vary from 8 to 76 years) in the literature for detailed and precise analysis of enjoyment smiles. On this database, we have verified the discriminative power of eyelid movement dynamics and showed its superiority over lip corner and cheek movements. We have evaluated fusion of features, and obtained minor improvements over using only eyelid features. We provided comparative results on three smile databases with the proposed method, as well as three other methods.

We report new and significant empirical findings on smile dynamics that can be leveraged to implement better facial expression analysis systems: 1) Maximum and mean apertures of eyes are smaller and speed of eyelid (also cheek and lip corner) movements are faster for *young people* compared to *adults*, during both spontaneous and posed smiles. 2) Mean eye aperture is smaller during spontaneous smiles in comparison to posed ones. 3) The speed and acceleration of eyelid movements are higher in posed smiles. 4) There is no significant difference in (movement) symmetry between spontaneous and posed smiles, even when tested separately for *young people* and *adults*.

8 Acknowledgments

This work is supported by Boğaziçi University project BAP-6531. Authors would like to thank Tomas Pfister for providing the executable of the CLBP-TOP implementation.

References

1. Ambadar, Z., Cohn, J.F., Reed, L.I.: All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. J Nonverbal Behav **33** (2009) 17–34
2. Ekman, P.: Telling lies: Cues to deceit in the marketplace, politics, and marriage. New York: WW. Norton & Company (1992)
3. Ekman, P., Hager, J.C., Friesen, W.V.: The symmetry of emotional and deliberate facial actions. Psychophysiology **18** (1981) 101–106
4. Ekman, P., Friesen, W.V.: Felt, false, and miserable smiles. J Nonverbal Behav **6** (1982) 238–252
5. Cohn, J.F., Schmidt, K.L.: The timing of facial motion in posed and spontaneous smiles. Int. Journal of Wavelets, Multiresolution and Information Processing **2** (2004) 121–132
6. Ekman, P., Friesen, W.V.: The Facial Action Coding System: A technique for the measurement of facial movement. Consulting Psychologists Press Inc., San Francisco, CA (1978)
7. Schmidt, K.L., Bhattacharya, S., Denlinger, R.: Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. J Nonverbal Behav **33** (2009) 35–45
8. Krumhuber, E.G., Manstead, A.S.R.: Can Duchenne smiles be feigned? New evidence on felt and false smiles. Emotion **9** (2009) 807–820
9. Manera, V., Giudice, M.D., Grandi, E., Colle, L.: Individual differences in the recognition of enjoyment smiles: No role for perceptual–attentional factors and autistic-like traits. Frontiers in Psychology **2** (2011)
10. Cohn, J.F., Reed, L.I., Moriyama, T., Xiao, J., Schmidt, K.L., Ambadar, Z.: Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. In: IEEE AFGR. (2004) 129–135
11. Valstar, M.F., Pantic, M.: How to distinguish posed from spontaneous smiles using geometric features. In: ACM ICMI. (2007) 38–45
12. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In: ICCV Workshops. (2011) 868–875
13. Dibeklioğlu, H., Valenti, R., Salah, A.A., Gevers, T.: Eyes do not lie: Spontaneous versus posed smiles. In: ACM Multimedia. (2010) 703–706
14. Tao, H., Huang, T.: Explanation-based facial motion tracking using a piecewise Bézier volume deformation model. In: CVPR. Number c (1999) 611–617
15. Schmidt, K.L., Cohn, J.F., Tian, Y.: Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. Biological Psychology **65** (2003) 49–66
16. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. on PAMI **27** (2005) 1226–1238
17. Valstar, M.F., Pantic, M.: Induced disgust, happiness and surprise: An addition to the MMI facial expression database. In: LREC, Workshop on EMOTION. (2010) 65–70
18. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., Wang, X.: A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Trans. on Multimedia **12** (2010) 682–691
19. Dibeklioğlu, H., Salah, A.A., Gevers, T.: A statistical method for 2-d facial landmarking. IEEE Trans. on Image Processing **21** (2012) 844–858
20. Schmidt, L.K., Ambadar, Z., Cohn, J.F., Reed, L.I.: Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. J Nonverbal Behav **30** (2006) 37–52
21. Dibeklioğlu, H., Gevers, T., Salah, A.A., Valenti, R.: A smile can reveal your age: Enabling facial dynamics in age estimation. In: ACM Multimedia. (2012)