EMOTIONAL SPEECH CLASSIFICATION USING GAUSSIAN MIXTURE MODELS AND THE SEQUENTIAL FLOATING FORWARD SELECTION ALGORITHM

Dimitrios Ververidis and Constantine Kotropoulos

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki 541 24, Greece
E-mail: {jimver, costas}@zeus.csd.auth.gr

ABSTRACT

Emotional speech classification can be treated as a supervised learning task where the statistical properties of emotional speech segments are the features and the emotional styles form the labels. The Akaike criterion is used for estimating automatically the number of Gaussian densities that model the probability density function of the emotional speech features. A procedure for reducing the computational burden of crossvalidation in sequential floating forward selection algorithm is proposed that applies the t-test on the probability of correct classification for the Bayes classifier designed for various feature sets. For the Bayes classifier, the sequential floating forward selection algorithm is found to yield a higher probability of correct classification by 3% than that of the sequential forward selection algorithm either taking into account the gender information or ignoring it. The experimental results indicate that the utterances from isolated words and sentences are more colored emotionally than those from paragraphs. Without taking into account the gender information, the probability of correct classification for the Bayes classifier admits a maximum when the probability density function of emotional speech features extracted from the aforementioned utterances is modeled as a mixture of 2 Gaussian densities.

1. INTRODUCTION

Emotion recognition is an area which attracts the interest of the research community [1]. A wide area of applications such as interface optimization and expressive voice synthesis are related to the classification of speech into emotional states. The decomposition of a probability density function (pdf) of multi-dimensional features into normal (Gaussian) densities is not a new idea [2]. The most reliable method for normal decomposition or Gaussian Mixture Modeling is the one invented by N. Day and latter by J. Wolfe at the late 60's [3] known as the *Expectation-Maximization (EM)* algorithm.

The EM algorithm was used to model the pdf of the emotional speech prsody features in [4, 5]. However, no special attention was paid to the estimation of the appropriate number of Gaussian densities in the mixture. In this paper, the *Sequential Floating Forward Selection* method (SFFS) [6] is employed to determine the best features among a set of 87 global statistical features with respect to the probability of correct classification for the Bayes classifier when the feature pdfs are modeled as mixtures of Gaussian densities. The work presented in this paper expands the already reported results in [7, 8]. In particular, the Akaike Information Criterion (AIC) [9] is now used for finding the appropriate number of Gaussian densities in the pdf of each feature given the emotional class. Moreover, the number of the crossvalidation repetitions is controlled through a *t-test* applied to the probability of correct classification for the Bayes classifier designed for various feature sets.

The outline of the paper is as follows. In Section 2, the data extracted from the Danish Emotional Speech (DES) database [10] are briefly described. In Section 3, feature extraction is presented. The features are scaled with techniques described in Section 4. The selection of an optimal set of features by SFFS is described in Section 5 and the discrimination capability offered by the Bayes classifier using such a feature set is assessed in Section 6. Finally, conclusions are drawn in Section 7.

2. DATA

The audio data used in the experiments consist of 1300 utterances, 800 more than those used in the previous works [7, 8], which is a typical database size. The utterances are manually extracted from the DES [10]. Each utterance is a speech segment between two silence pauses. The 800 utterances that are now added in dataset are detached from paragraphs whereas the previously employed 500 utterances were associated to isolated words and sentences. All utterances are expressed by 4 professional actors (2 males and 2 females) in 5 emotional styles such anger, happiness, sadness, surprise, and neutral style. The extended set provides the necessary amount of data to train a GMM (Gaussian Mixture Model) with up to 3 Gaussian densities. An end-point detection algorithm was used in order to find the start and end of an utterance. Explosive fricatives, which often arise at the tails of the words, have been excluded, because they do not reveal any information on the emotional content of speech [11].

3. FEATURE EXTRACTION

The so-called global statistical short-term features [12], i.e., statistical properties of formant, pitch, and energy contours of the speech signal are used. The short-term features are estimated on a frame basis, $f_s(n;m) = s(n)w(m-n)$, where $s(n)$ is the speech signal and $w(m-n)$ is a window of length N_w ending at sample m. For example, the short-term energy of the speech frame ending at m is

$$E_s(m) = \frac{1}{N_w} \sum_{n=m-N_w+1}^{m} |f_s(n;m)|^2. \tag{1}$$

The work presented was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

In order to estimate the pitch, the signal is low filtered at 900 Hz and then "center clipping" is applied to each frame [13]. The procedure is as follows:

$$\hat{f}_s(n;m) = \begin{cases} f_s(n;m) - C & \text{if } |f_s(n;m)| > C \\ 0 & \text{if } |f_s(n;m)| < C \end{cases} \quad \forall n \qquad (2)$$

where C is set at the 30% of the maximum value of $f_s(n;m)$. Clipping is a non-linear procedure that prevents the 1^{st} formant from interfering with pitch. The pitch frequency is estimated by the short-term autocorrelation

$$r_s(\eta;m) = \frac{1}{N_w} \sum_{n=m-N_w+1}^{m} \hat{f}_s(n;m)\hat{f}_s(n-\eta;m) \qquad (3)$$

where η is the lag. Let us assume that the pitch frequency is limited in the range $[60, 320]$ Hz. The pitch frequency of the frame ending at m is given by

$$P_s(m) = \frac{F_s}{N_w} \operatorname{argmax}_\eta \{|r(\eta;m)|\}_{\eta=N_w\ (F_l/F_s)}^{\eta=N_w\ (F_h/F_s)} \qquad (4)$$

where F_s is the sampling frequency, $F_l = 60$ Hz, and $F_h = 320$ Hz.

To estimate the 4 formants, we find the angle of the poles for the all-pole vocal tract model

$$\hat{\Theta}(z) = \frac{1}{1 - \sum_{i=1}^{M} \hat{a}(i) z^{-i}} \qquad (5)$$

in the z-plane and consider the poles that are further from zero as indicators of formant frequencies. In (5), $\hat{a}(i)$ are estimated by the Levinson-Durbin algorithm [13] and M is the order of the model, which is usually selected as 10-12 for speech sampled at 8 kHz.

Many global statistical features (e.g., the mean speech energy) admit a single value throughout the entire utterance. For such global statistical features, the time information is lost. In order to cope with this loss, we include also features related to the duration of the rising and falling slopes of the pitch and energy contours. After speech framing, we estimate the pitch, the energy, and the formants of each speech frame, as was previously explained. In order to create a contour for each feature, we assign the feature value computed on a frame basis to all samples belonging to the frame. For example, the energy contour is given by

$$e(n) = E_s(m), \quad n = m - N_w + 1, \ldots, m \qquad (6)$$

where $E_s(m)$ is the short term energy of the frame $f(n;m)$. To determine which samples belong to a set of rising slopes (S_r), falling slopes (S_f), plateaux at maxima (S_{ma}), and plateaux at minima (S_{mi}), we estimate the first derivative of the feature contour by numerical methods. For example, the derivative of the energy contour can be estimated by the first-order difference $e_D(n) = e(n) - e(n-1)$, $n = 2, \ldots, L$, where L is the signal length. Subsequently, the algorithm of Figure 1 is applied. In this algorithm, a is a constant that enables the detection of the rising or falling slopes and the plateaux. The distinction between the plateaux at maxima and those at minima is accomplished with the constant b which is set to 0.45. The same set of 87 global statistical features used in [7, 8] is employed in this paper as well.

```
if $e_D(n) \geq a$, $s(n) \in S_r$
else if $e_D(n) \leq -a$, $s(n) \in S_f$
else if $|e_D(n)| < a$
    if $s(n) > \max(s(i)) \cdot b$, $s(n) \in S_{ma}$
    else if $s(n) \leq \max(s(i)) \cdot b$, $s(n) \in S_{mi}$
    end
end
```

Fig. 1. Algorithm for finding the plateaux at minima/maxima and the rising/falling slopes of pitch and energy contours.

4. DATA PREPROCESSING

The extracted features undergo a preprocessing, because it was found that a bad scaling can cause overflow or underflow errors during the estimation of the covariance matrices $\mathbf{\Sigma}_i$, endless loops in the EM algorithm, and it has a biased influence in the visual representation after dimensionality reduction with Principal Component Analysis. The preprocessing consists of two steps, namely the normalization and the handling of missing data.

Each feature X_k, $k = 1, \ldots, 87$, has its own dynamic range. Features with variance of order 10^6 such as the fourth formant, have greater influence in the classifier design than features with a variance of order 10^2 such as the mean value of pitch. Thus, a linear transformation is applied to each feature. If $a_k = \min_i\{X_{ki}\}$ and $b_k = \max_i\{X_{ki}\}$ for $i = 1, ..., N_s$ where N_s equals the total number of utterances, then the linear transformation from $[a_k, b_k] \rightarrow [0, 1]$ is defined as $\hat{X}_{ki} = \frac{X_{ki} - a_k}{b_k - a_k}$ $i = 1, \ldots, N_s$. Since the pdf of many features X_k is not evenly distributed about the mean and it has the shape of an exponential distribution, we avoid the application of the whitening normalization, because the exponential distribution is not symmetrical and the outliers of the exponential distribution will be moved further away. The exponentially distributed features may lead to an increased computational time and underflow warnings, as they become too dense near the lower bound which in our case is 0_+. Accordingly for exponentially distributed features such as those indexed by {1, 14, 22, 31-33, 35, 39-41, 46-48, 55-58, 66-68, 70-73, 78, 79, 81, 85-87} [7], after the linear transformation, we apply the transformation $\hat{\hat{X}}_k = \frac{1 - e^{-\lambda \hat{X}_k}}{1 - e^{-\lambda}}$ where λ is set to its maximum likelihood estimator $1/E\{\hat{X}_k\}$.

There are cases where X_k can not be estimated at utterance i. For example, some pitch contours do not have a plateaux below the 0.45% of the maximum pitch value. When there is a large number of missing data, the corresponding feature X_k is discarded. Features such as those indexed by $\{9, 24 - 30, 34, 38, 42, 49, 59 - 65, 69, 77, 84\}$ are discarded, because their missing data ratio varies between 2% and 70%. In the cases where the missing data are less than 1% of the whole feature data, the missing data are replaced with the sample mean. Proceeding so, only 65 among the 87 features are retained.

5. SEQUENTIAL FLOATING FORWARD SELECTION ALGORITHM

In the following, we omit the hats from features for notation simplicity. Let $\mathcal{X}$ denote the feature set, i.e., $X_k \in \mathcal{X}$. After each forward step, the SFFS algorithm [6] applies a number of backward steps as long as the resulting subsets are better than the previously derived ones at this level. Consequently, there are no backward steps at all when the performance cannot be improved. Starting form an initial empty set of features $\mathcal{Z}_0$, at each inclusion step at the level l we seek the feature $X^+ \in (\mathcal{X} - \mathcal{Z}_{l-1})$ such that for $\mathcal{Z}_l = \mathcal{Z}_{l-1} \cup \{X^+\}$ the probability of correct classification of the Bayes classifier $J(\mathcal{Z}_l)$ is maximized. The inclusion step is followed by a conditional exclusion step. We exclude at level l those $Z^- \in \mathcal{Z}_l$ as long as the correct classification of the Bayes classifier for the feature set $\mathcal{Z}_{l-1} = \mathcal{Z}_l - Z^-$, $J(\mathcal{Z}_{l-1})$, is higher than $J(\mathcal{Z}_l)$.

In the following, we describe the Bayes classifier design, the estimation of probability density functions by Gaussian mixtures, the estimation of the number of Gaussian densities, and finally the estimation of the probability of correct classification of the Bayes classifier by crossvalidation for $nrep$ repetitions.

Let the feature X_k be treated as the kth element of a random vector $\mathbf{x}$ (e.g. a pattern). Let ω_i denote the ith class, $P(\omega_i)$ be the a priori probability of class ω_i, and $p(\mathbf{x}|\omega_i)$ be the class conditional pdf. The Bayes classifier assigns the pattern $\mathbf{x}$ to ω_i if

$$P(\omega_i)\; p(\mathbf{x}|\omega_i) > P(\omega_j)\; p(\mathbf{x}|\omega_j) \quad j = 1, 2, \ldots, c \tag{7}$$

where c is the total number of classes. In our case, $c = 5$. Let $\mathcal{L}_i$ be the region where $\mathbf{x}$ is classified to ω_i and $\mathcal{L} = \cup_{i=1}^{c} \mathcal{L}_i$. We also define the complement of $\mathcal{L}_i$ as $\mathcal{L}_i^c = \mathcal{L} - \mathcal{L}_i$. The probability of error of the Bayes classifier is given by

$$\varepsilon = \sum_{i=1}^{c} P(\omega_i) \int_{\mathcal{L}_i^c} p(\mathbf{x}|\omega_i)\; d\mathbf{x}. \tag{8}$$

The pdf $p(\mathbf{x}|\omega_i)$, where $\mathbf{x}$ is a d-dimensional pattern can be modeled by a mixture of Gaussian densities, i.e.

$$p(\mathbf{x}|\omega_i) \;=\; p(\mathbf{x}|\boldsymbol{\Theta}_i) = \sum_{j=1}^{N_{im}} \pi_j \; g(\mathbf{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \tag{9}$$

$$g(\mathbf{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \;=\; \frac{\exp[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j)]}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_j|^{1/2}} \tag{10}$$

where the parameter $\boldsymbol{\Theta}_i = \{\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{N_{im}}$ consists of the mixture weight π_j, the mean vector $\boldsymbol{\mu_j}$, and the covariance matrix $\boldsymbol{\Sigma}_j$ of the jth Gaussian component for $j = 1, 2, \ldots, N_{im}$, and N_{im} is number of mixtures in the ith class. $\boldsymbol{\Theta}_i$ can be estimated by applying the EM algorithm [3]. Let $\{\mathbf{x}^t\}_{t=1}^{t=N_{\mathcal{D}}}$ be the training samples we have at our disposal. The EM algorithm finds

$$\boldsymbol{\Theta}_i^* = \arg\max \prod_{t=1}^{N_{\mathcal{D}}} P(\mathbf{x}^t | \boldsymbol{\Theta}_i). \tag{11}$$

The number of Gaussian densities in (9) can be estimated by applying the Akaike Information Criterion (AIC) [9] or the Minimum Description Length principle which evaluates the parsimony of the model when N_{im} components are employed. The AIC is given by

$$AIC_{N_{im}} = -\ell(\boldsymbol{\Theta}_i^* \mid \mathbf{x}) + 2K \tag{12}$$

where $\ell(\boldsymbol{\Theta}_i^* \mid \mathbf{x})$ is the maximum log-likelihood and K is a penalty on the complexity that depends on the number of the free parameters, i.e.

$$K = N_{im}(1 + d + \frac{d}{2}(1 + d)). \tag{13}$$

An order-recursive procedure can be applied for the optimal estimation of N_{im} and $\boldsymbol{\Theta}^*_{i|N_{im}}$. We increase the number of Gaussian densities by 1, we estimate $\boldsymbol{\Theta}^*_{i|N_{im}+1}$ by applying the EM-algorithm and we calculate $AIC_{N_{im}+1}$ from (12-13). We stop increasing the model order if

$$AIC_{N_{im}+1} - AIC_{N_{im}} > 0. \tag{14}$$

Let

$$J_{nrep}(\mathcal{Z}) = 1 - E\{\varepsilon(\mathcal{Z}, \mathcal{T}_r; \mathcal{D}_r)\} \tag{15}$$

where $\varepsilon(\mathcal{Z}, \mathcal{T}_r; \mathcal{D}_r)$ is the probability of error for the Bayes classifier that was designed using $\mathcal{D}_r$ during the training when it is applied to $\mathcal{T}_r$ and the expectation $E\{\}$ is applied over the sequence of error probabilities measured over $\mathcal{T}_r, r = 1, 2, \ldots, nrep$. In (15), the dependence of J_{nrep} on the feature set $\mathcal{Z}$ is made explicit. 90% of the available utterances are used to build $\mathcal{D}_r$ and the remaining 10% creates $\mathcal{T}_r$.

The computation of a fixed large number of crossvalidation repetitions (e.g. $nrep > 30$) adds frequently an unnecessary computational burden. A novel method for estimating the best feature at each forward and backward iteration of the SFFS is proposed that does not resort to a large number of crossvalidation repetitions. We assume that the probability of error of the Bayes classifier for a large number of crossvalidation repetitions follows a normal distribution. The goal at level l is to find which non-selected feature $X \in (\mathcal{X} - (\mathcal{Z}_{l-1}))$ yields the greatest improvement in the probability of correct classification for the Bayes classifier among the non-selected features

$$J_{max} = \max_{X \in (\mathcal{X} - \mathcal{Z}_{l-1})} J_{nrep}(\mathcal{Z}_{l-1} + \{X\}). \tag{16}$$

If $nrep$ is a large number, then $J_{nrep}(\mathcal{Z}_{l-1} + \{X\})$ is an accurate estimate of the maximum probability of correct classification one might expect from the Bayes classifier. But the computation is time consuming. If $nrep$ is small, $J_{nrep}(\mathcal{Z}_{l-1} + \{X\})$ is computed faster, but it is not accurate. Let us separate the features $X \in (\mathcal{X} - \mathcal{Z}_{l-1})$ in potentially expressive features and potentially bad features. The former features yield $J_{nrep1}(\mathcal{Z}_{l-1} + X) \geq J(\mathcal{Z}_{l-1})$, while the latter ones consistently yield $J_{nrep2}(\mathcal{Z}_{l-1} + X) < J(\mathcal{Z}_{l-1})$. We propose to formulate a t-test in order to test the hypothesis $J_{nrep2}(\mathcal{Z}_{l-1} + X) < J(\mathcal{Z}_{l-1})$ at 95% significance level for a small number of iterations (e.g. $nrep2$=10). If the hypothesis is accepted, we discard the feature X. Otherwise, we perform more iterations. If the hypothesis $J_{nrep1}(\mathcal{Z}_{l-1} + X) \geq J(\mathcal{Z}_{l-1})$ is accepted at 95% significance level for $10 < nrep1 \leq 50$, we include the feature under study in the set of 10 best features.

6. RESULTS

Two experiments were conducted using 500 and 1300 utterances. The set of 500 utterances contains speech segments from isolated words and sentences. In the following, let us call it set $\mathcal{A}$. Let $\mathcal{B}$ be the set of 1300 utterances that contains speech segments from words, sentences, and paragraphs. From the inspection of Table 1 we conclude that set $\mathcal{A}$ has a stronger emotional valence than the set $\mathcal{B}$, because the corresponding probability of correct classification for the Bayes classifier on set $\mathcal{A}$ is 6-10% higher than that of the Bayes classifier on set $\mathcal{B}$. This implies that the emotional valence of paragraphs is lower than the valence of isolated words and sentences.

Table 1. Probability of correct classification.

Gender	Number of Gaussian densities					Human rates
	Set $\mathcal{A}$		Set $\mathcal{B}$			
	1	2	1	2	3	
Both	0.542	0.562	0.485	0.480	0.464	0.673
Male	0.660	0.625	0.560	0.540	0.503	0.676
Female	0.600	0.540	0.509	0.487	0.485	0.669

A GMM with more than 1 Gaussian densities improves the probability of correct classification of the Bayes classifier only in the set $\mathcal{A}$ when the male and female utterances are not discriminated. When the emotional valence is low, as in set $\mathcal{B}$, no gain was obtained by attempting to train separate emotional speech classifiers for males and females.

If we model all emotional style pdfs by a single Gaussian density and we build separate emotional speech classifiers for male utterances and female ones we obtain significant gains in the probability of the correct classification. That is, a improvement of 12% was obtained for male utterances and of 6% for females. Attempting to model the emotional style pdfs with GMMs of 2 Gaussian densities improvements were still obtained, but they are inferior to those obtained with a single Gaussian. In Table 2, the best combination of 10 features for each experiment is indicated. The energy below 250 Hz (index 1) is present in all combinations. The minimum value of the first formant (index 10) is also quite frequent. The mean value of energy within the rising slopes of the energy contours (index 78) is found to be also important. The SFFS algorithm overrides the local minima during the search of the best combination of features. The comparison with the previously reported results [7], where the Sequential Forward Selection (SFS) algorithm was employed, reveals that the SFFS algorithm improves the probability of correct classification of the Bayes classifier by 3% in all the corresponding categories.

Table 2. Best combination of 10 features selected by the Sequential Floating Forward Selection algorithm in set $\mathcal{A}$.

Classifier	Best feature combination
Both genders	
1 Gaussian	1, 3, 6, 10, 32, 55, 56, 68, 74, 80
AIC-GMM 2	1, 4, 10, 12, 20, 22, 32, 44, 67, 78
Males only	
1 Gaussian	1, 8, 10, 20, 32, 43, 54, 56, 78, 80
AIC-GMM 2	1, 8, 20, 47, 57, 58, 78
Females only	
1 Gaussian	1, 4, 16, 21, 22, 39, 54, 67, 74, 78
AIC-GMM 2	1, 20, 43, 78

7. CONCLUSIONS

We have demonstrated that the SFFS algorithm is more powerful for feature selection than the SFS one when emotional speech Bayes classifiers are built either by taking into account or ignoring the gender information. We have also proposed and tested a novel procedure for reducing the computational burden of crossvalidation in SFFS algorithm that applies the t-test on the probability of correct classification for the Bayes classifier designed for various feature sets. The highest probability of correct classification for the Bayes classifier has been obtained working with the set of utterances from isolated words and sentences when separate emotional speech classifiers were built for male and female utterances and the speech emotional style pdfs were modeled with a single Gaussian density.

8. REFERENCES

[1] K. R. Scherer, "Emotion effects on voice and speech: Paradigms and approaches to evaluation," in *Proc. ISCA Workshop on Speech and Emotion 2000 (ITWR)*, Belfast, 2000, vol. 1, Invited review paper.

[2] K. Fugunaka, *Introduction to Statistical Pattern Recognition*, N.Y. Academic Press:, 1990.

[3] J. H. Wolfe, "Pattern clustering by multivariate analysis," *Multivar. Behav. Res.*, vol. 5, pp. 329–359, 1970.

[4] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - belief network architecture," in *Proc. 2004 IEEE Int. Conf. Acoustics, Audio and Signal Processing*, May 2004, vol. 1, pp. 577–580.

[5] C. M. Lee and S. Narayanan, "Towards detecting emotion in spoken dialogs," *IEEE Trans. Speech & Audio Processing*, vol. 13, no. 2, pp. 293–303, 2005.

[6] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. 12th IAPR Int. Conf. Pattern Recognition*, Israel, 1994, vol. 1, pp. 279–283.

[7] D. Ververidis and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information," in *Proc. 12th European Signal Processing Conf.*, Austria, September 2004, pp. 341–344.

[8] D. Ververidis and C. Kotropoulos, "Emotional speech classification using gaussian mixture models," in *Proc. IEEE Int. Symp. Circuits and Systems*, Japan, 2005, to appear.

[9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, vol. 19, no. 6, pp. 716–723, December 1974.

[10] I. S. Engberg and A. V. Hansen, *Documentation of the Danish Emotional Speech (DES) Database, Internal AAU report*, Center for Person Kommunikation, Denmark, 1996.

[11] B. D. Womack and J. H. L. Hansen, "N-channel Hidden Markov Models for combined stressed speech classification and recognition," *IEEE Trans. Speech & Audio Processing*, vol. 7, no. 6, pp. 668–677, 1999.

[12] R. Cowie and E. Douglas-Cowie, "Automatic statistical analysis of the signal and prosodic signs of emotion in speech," in *Proc. Fourth Int. Conf. Spoken Language Processing*, USA, October 1996, pp. 1989–1992.

[13] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discete-Time Processing of Speech Signals*, N.Y.: Wiley, 2000.