COMPARISON OF NOISE ROBUST METHODS IN LARGE VOCABULARY SPEECH RECOGNITION

Sami Keronen[1], *Ulpu Remes*[1], *Kalle J. Palomäki*[1], *Tuomas Virtanen*[2], *and Mikko Kurimo*[1]

[1] Adaptive Informatics Research Centre, Aalto University
P.O. Box 15400, FI-00076 Aalto, Finland
phone: + (358) 947025388, email: firstname.lastname@hut.fi, web: www.cis.hut.fi/projects/speech/
[2] Department of Signal Processing, Tampere University of Technology
P.O. Box 553, FI-33101 Tampere, Finland
phone: + (358) 331154798, email: firstname.lastname@tut.fi, web: http://arg.cs.tut.fi/

ABSTRACT

In this paper, a comparison of three fundamentally different noise robust approaches is carried out. The recognition performances of multicondition training, Data-driven Parallel Model Combination (DPMC), and cluster-based missing data reconstruction methods implemented in a large vocabulary continuous speech recognition system are evaluated with Finnish language speech data consisting of real recordings in noisy environments. All three methods improve the recognition accuracy substantially in poor signal-to-noise ratio (SNR) conditions when compared to a baseline system trained on clean speech. DPMC and missing data reconstruction systems give the best performance on high SNR conditions. On low SNR conditions, the performance of multicondition trained system is ranked the best, DPMC the second best and missing data reconstruction the third.

1. INTRODUCTION

In automatic speech recognition (ASR), noise robustness may be addressed by several fundamentally different methods. One method is to train the system directly on a specific type of noise encountered in the recognition phase. This type of system is called a matched system and it is likely to be superior compared to any noise compensation method, but only for that specific type of noise. Adjusting the system for new types of noises requires a large database of new noise types and time consuming re-training of the system. A more practical alternative to the matched training is multicondition training, in which the system is trained directly on noisy speech encountered in the most common noise environments, thus preventing the need for re-training the system every time the background noise changes. Re-training of a multicondition trained system is only required if the noise encountered is considerably different from the noises included in the training set. A multicondition trained system is evaluated in this study.

It is possible to overcome the requirements of massive data collection of different acoustic environments as well as the requirement of re-training for each new type of noise by using active noise compensation methods. One active compensation method is to compensate for the noise in the feature extraction stage to obtain an estimate of the underlying clean speech signal from the noise contaminated speech signal. This can be achieved through pre-processing techniques such as Vector Taylor Series (VTS) expansion described in [11]. Another active compensation method is the model compensation. The idea behind model compensation is not to pre-process the signal but rather allow the presence of noise in the recognition process by adapting the acoustic model to match the current noisy environment. Model compensation methods include, for example, Parallel Model Combination (PMC) [7] and Noise Adaptive Training [10]. In PMC, the parameters of speech and noise models such as Mel-frequency cepstral coefficients (MFCCs), are transformed from the cepstral domain to linear spectral domain (where speech and noise are assumed additive), combined using a mismatch function and transformed back to cepstral domain. Since it is not possible to calculate the mismatch functions in closed form, Data-driven PMC, the second applied approach in this study, can be used for accurate estimations [7].

Finally, there is a class of methods based on finding reliable information in the observed features. Missing data methods [14], for example, divide the noisy observations to reliable and unreliable spectrotemporal components depending on whether the features are dominated by speech or noise — an approach motivated by studies on auditory scene analysis (ASA) which demonstrate that the most intense sound component in each auditory scene dominates the combined neural response to the scene [2]. Thus, the speech dominated components may be regarded as reliable estimates for the underlying clean speech and used as such in speech recognition, while the noise dominated components are taken to represent only noise, and the speech information carried in these components is considered missing. The missing information can, however, be reconstructed using e.g. cluster-based reconstruction [15].

In this paper, a comparison of three fundamentally different noise robust ASR approaches including multicondition training, DPMC, and cluster-based missing data reconstruction is carried out. The three noise robust approaches are implemented in a large vocabulary continuous speech recognition system and evaluated on real noisy speech recordings in car and public place environments. As far as the authors are aware, no direct comparison between these three approaches has been carried out previously nor has DPMC been evaluated on real noisy recordings. In previous studies on DPMC e.g. [5], [7] and PMC e.g. [12], speech is usually corrupted with artificially added noise. Such experiments are slightly unrealistic since there is no convolutive noise (e.g. reverberation) that may affect the DPMC performance.

2. METHODS

This section describes the baseline system (Sect. 2.1) and the three noise robust approaches based on multicondition training (Sect. 2.2), DPMC (Sect. 2.3.2) and cluster-based missing data reconstruction (Sect. 2.4). A voice activity detector (Sect. 2.5) used with both DPMC and missing data reconstruction is also described.

2.1 Baseline system

The baseline system used in this work is a large vocabulary continuous speech recognizer based on hidden Markov models (HMM) with state likelihoods modeled by Gaussian mixture models (GMM). The baseline system uses a morph-based variable length n-gram language model [18] trained on 145 million words of book and newspaper data. The decoding vocabulary is practically unlimited [8] as all words and word forms can be represented with the unsupervised morphs. The language model is combined to the acoustic model using scaling factor on the language model log-probability. The scaling factor is optimized for noiseless speech with respect to the letter error rate (LER). The decoder is a time-synchronous beam-pruned Viterbi token-pass system and the acoustic models are state-tied tri-phones constructed with a decision-tree method. Each state is modeled with a maximum of 100 Gaussians and the states are associated with gamma probability functions to model the state durations [13]. The speech signal is represented with 13 MFCC features concatenated with their first and second order differentials, scaled, and mapped with maximum likelihood linear transformation (MLLT) [6] optimized in training. The width of the dynamic parameter window is 2 frames (512 samples) and the dynamic parameters are calculated according to DPMC requirements as simple differences over the given window width. Finally, the covariance matrix of each Gaussian is diagonalized.

2.2 Multicondition training

In multicondition training, the system is trained on speech in varying noise conditions, which are ideally the ones that are most likely encountered in the use of recognition system. Multicondition training does not increase the computational complexity as it does not require any changes to an existing ASR system; only the training material is changed. Obtaining a sufficient amount of training data representing the variation of noise conditions well enough is the main challenge of multicondition training. However, multiple noise conditions can also be simulated by artificially adding noise with varying characteristics and SNR to the training speech data.

The multicondition trained system used in this work was trained on data consisting of 50 % noiseless speech, 30 % speech recorded in public places, and 20 % speech recorded in moving cars. Three SNR levels (microphone positions) and four different microphones were included in the noisy training data. A more thorough description of the training set is given in Section 3. The number of Gaussians in the multicondition system was approximately the same as the baseline system after state tying executed during the training process.

2.3 Parallel Model Combination

PMC [7] is a model compensation technique used to adapt the acoustic model trained on noiseless speech to the current recognition environment. Standard PMC assumes the feature vectors are linear transforms of the log spectrum such as MFCCs, and that the noise is primarily additive. First the Gaussian means and covariances of the speech and noise models are mapped from the cepstral domain to the linear spectral domain where the noise is assumed additive. Then the means and covariances of the combined speech and noise distributions are computed and transformed back to the cepstral domain. When the distribution of the original features is modeled with a GMM, the distributions of the resulting transformed features do not have a simple functional form. Despite this, they are usually approximated with a GMM.

Different mapping approximations have been developed to the nonlinear transformation between log-spectral domain and linear spectral domain. For example, log-normal and log-add [7] approximations are simple and fast methods but give rather poor mapping approximations resulting in large recognition error rates, especially on low SNR conditions, because they can approximate only the transformations of static parameters. A more accurate mapping is achieved by combining the transformed parameters using mismatch functions [7]. However, it is not possible to solve the mismatch functions in closed-form. Solution for this problem is addressed in the next section.

2.3.1 Data-driven PMC

DPMC [7] is a formulation of PMC developed to circumvent the unsolvability of the mismatch functions. DPMC uses the distributions of the clean speech and noise parameters to generate virtual data points in a Monte Carlo process. Data for static and dynamic parameters (differential features, see 2.1) are generated independently. Then the virtual data points are combined with the expectation values of mismatch functions, the static and dynamic parameters separately, to form virtual noisy speech data points. The mean and covariance are calculated from the virtual noisy speech data points, which yields an estimate of the noisy speech distribution. Then the combined distribution parameters are mapped back to cepstral domain. A more thorough description of mismatch functions can be found in [7, pp. 38–39]. The accuracy of noisy speech distribution estimate depends on the number of generated data points. However, as the number of data points increases, the computational cost grows linearly, which accumulates with the increased computational intensity on large vocabulary systems, in which large speech models are used.

2.3.2 DPMC system

DPMC system applied in this work used a standard diagonal covariance approximation described in [7, pp. 55–57]). The noise model was estimated on-line separately for each utterance from speech pauses within the corresponding utterance. A voice activity detector, described in Section 2.5, was used to find speech pauses. For the combination process, 50 data points per Gaussian were generated. The computational complexity of DPMC implementation is $O(G(M^3+P*M^2))$ per combination, where G is the number of Gaussians in the acoustic model, M is the number of Mel bands, and P is the number of generated data points.

2.4 Missing data approach

When speech is corrupted with additive noise, the log-Mel-representation of the noisy speech signal is divided into reliable i.e. speech dominated regions and unreliable i.e. noise dominated regions where speech information is effectively missing. Thus, the missing feature approaches used for noise compensation in ASR need first to identify the unreliable feature components and then handle speech recognition with partially observed data. Motivation for the missing data approach originally comes from human speech perception and auditory scene analysis (ASA) [1].

2.4.1 Noise mask estimation

In this work, the unreliable feature components are identified based on local SNR estimates in the Mel-spectral domain. These are derived using a stationary noise estimate which is calculated from non-speech frames identified using the voice activity detector described in Section 2.5. Noisy observations are then considered reliable if the observed value exceeds the estimated noise power with a predefined threshold $\gamma = 3$ dB and unreliable otherwise. This provides an initial noise mask estimate. Based on the glimpse model proposed in [3], the mask is then processed to contain only connected regions of reliable speech features that are larger than a certain minimum size, i.e. glimpses of speech. The decision to include only glimpses larger than a minimum size five and to use threshold value $\gamma = 3$ dB in the mask estimation is based on experiments with the noisy development data (see Section 3 for dataset description).

2.4.2 Missing data reconstruction

The missing data methods used in ASR may be divided into classifier modification and data reconstruction approaches. In classifier modification, the speech recognition system is modified to handle the unreliable features using e.g. bounded marginalization [4], while reconstruction is based on replacing the missing values with clean speech estimates in the log-compressed Mel-spectral domain. The reconstructed features can be transformed to an arbitrary feature domain and subjected to normalization and adaptation, which makes this a practical approach also for large vocabulary continuous speech recognition systems that typically use such transformations.

In this work, the missing values are reconstructed using the cluster-based missing data reconstruction method proposed in [15]. The method calculates a bounded maximum likelihood (BMAP) estimate for the missing values based on the observed features and a priori information on the statistical dependencies between spectral components in clean speech. The features are reconstructed in the log-Mel-spectral domain and mapped to the acoustic model feature domain described in Section 2.1 after reconstruction. The clean speech model used in this work is a 5-component GMM trained with a 52-minute dataset of read sentences randomly extracted from the SPEECON training data (Section 3). The clusters and model parameters were jointly estimated using the expectation-maximization (EM) algorithm in GMMBAYES Matlab toolbox. For a detailed description on solving the BMAP estimates in cluster-based reconstruction, see [14]. The computational complexity of the missing data reconstruction implementation is $O(F * C * M^4)$ per utterance, where F is the number of frames in the utterance, C is the number of components in the reconstruction model, and M is the number of Mel-frequency bands.

2.5 Voice Activity Detector

The voice activity detector used in DPMC and missing data systems was based on an HMM/GMM speech/non-speech classifier which is presented and examined in more detail in [16]. The current version differs from the previous only by newly trained speech and non-speech models. To ensure classification of speech and noise together as speech, the training material in the current version contained broadcast news with artificially added noise instead of plain TV news material. This lead to improved frame classification accuracy of 98 % for close microphone and 94 % for far microphone on public place data (93 % and 92 % in previous version, respectively) in the same test described fully in [16].

3. DATA

All data used in this work was taken from SPEECON [9] Finnish language corpus which contains both spontaneous and read speech. Two equally sized training sets were constructed both comprising of 293 speakers and containing approximately 19.5 hours of speech. The first training set, referred to as the clean training set, was used to train the baseline system and contained speech recorded in sufficiently noiseless environments (estimated average SNR of 26 dB according to SPEECON database specifications). The second training set was used in multicondition training and it consisted of half noiseless speech and half noisy recordings from the public place and car environments. Only the training sets contained spontaneous speech. A development set was also constructed from SPEECON database and it consisted of 1.2 hours of noiseless speech from 39 speakers. The development set was used for language model scaling optimizations.

The systems were evaluated with noisy data selected from public and car environments. The noisy car test set consisted of 57 minutes of speech from 20 speakers, and public place test set consisted of 94 minutes of speech from 30 speakers. The corresponding development set lengths were 29 and 60 minutes. The training, development, and evaluation sets were exclusive i.e. they do not share speakers. Three microphone position recordings are used in the evaluation set. The closest microphone was a headset which had an estimated average SNR of 13 dB in car and 24 dB in public place recordings. The second closest microphone was attached to the chest level and had an estimated average SNR of 5 dB in the car and 14 dB in public place recordings. The farthest microphone in the car recordings was located in the rear-view mirror approximately one meter away from the speaker and had an estimated average SNR of 8 dB. The farthest microphone in the public place recordings was placed 0.5–1 meter away from the speaker and had an estimated average SNR of 9 dB.

The inconsistency between the SNRs of the mid and far distance microphones in the car recordings is caused by the different characteristics of the microphones. The far microphone in the car recordings (AKG Q400 Mk3T) has been designed to be a part of a hands-free system and therefore it has a limited frequency response in order to make the microphone robust to e.g. engine and turbulence noises. The estimated average SNR for this microphone is higher relative to the mid distance microphone as it does not pick the high intensity low frequency components of the engine noise due to its restricted bandwidth.

Table 1: Public place noise evaluation set error rates (LER/WER %).

Mic	Baseline	DPMC	Multi-condition	Missing data
Close	5.7/20.2	4.3/16.2	7.5/26.8	4.5/16.6
Mid	38.4/63.9	15.7/32.5	9.4/25.1	24.9/54.1
Far	54.6/76.5	28.3/44.6	17.7/35.1	38.0/66.3

Table 2: Car noise evaluation set error rates (LER/WER %).

Mic	Baseline	DPMC	Multi-condition	Missing data
Close	6.7/21.8	5.2/17.1	5.9/19.6	4.9/16.9
Mid	64.3/86.3	29.2/48.1	13.0/33.3	32.6/54.4
Far	87.6/99.5	79.3/94.3	42.5/72.0	75.7/97.5

4. RESULTS

Letter (LER) and word error rates (WER) on the public place noise evaluation set are collected in Table 1. LER is the common measurement unit used in Finnish speech recognition. The baseline system performance shows the largest degradation with decreasing SNR, from the close recordings (5.7 %) to mid (38.4 %) and far (54.6 %) recordings. DPMC system is the most accurate system on close recordings (4.3 %) and the second most accurate on mid (15.7 %) and far recordings (28.3 %). Multicondition trained system gives the best performance on mid (9.4 %) and far (17.7 %) recordings by a large margin but performs worse than the baseline system on close recordings (7.5 %). Missing data system performs better than the baseline system in all conditions (4.5 %, 24.9 %, and 38.0 %) but is poorer than multicondition trained or DPMC systems in noisier mid and far recordings.

Letter and word error rates on car noise evaluation set are collected in Table 2. Again the baseline system shows the largest degradation in recognition performance with decreasing SNR from close recordings (6.7 %) to mid (64.3 %) and far (87.6 %) recordings. DPMC (5.2 %) and multicondition systems (5.9 %) perform better than the baseline system on close recordings but the lowest error rate is achieved by the missing data system (4.9 %). DPMC has the worst performance on far (79.3 %) recordings and the second worst on mid (29.2 %) recordings. The multicondition system has the best performance on mid (13.0 %) and far (42.5 %) recordings. The missing data system performance on mid (32.6 %) recordings is slightly lower than the respective DPMC performance and on far (75.7 %) recordings, it is slightly higher than the respective DPMC performance.

Wilcoxon signed rank test was used for pairwise statistical comparisons between the letter error rates of each system on combined public place and car noise. The Z-scores are collected in Table 3. Based on a 95 % confidence interval, the Z-score greater than 1.96 or less than -1.96 indicates a statistically significant difference in the comparison. Statistically significant difference was reached in all other pairwise comparisons except in between the baseline-multicondition and DPMC-missing data systems on close recordings.

Table 3: Z-scores of pairwise comparisons and the respective relative LER improvements (left relative to right) on combined public place and car noises (z-score/relative LER improvement %). Bold numbers indicate a statistical difference between the systems.

Pair*	Close	Mid	Far
BL-DPMC	**-3.68/24.6**	**-6.15/56.9**	**-5.67/28.4**
BL-MC	-1.95/-13.1	**-6.10/77.8**	**-6.50/59.4**
BL-MD	**-5.60/24.6**	**-6.15/42.5**	**-6.15/21.9**
DPMC-MC	**-4.95/-50.0**	**-5.84/48.6**	**-5.32/43.3**
DPMC-MD	-0.94/0.0	**-5.24/-33.3**	**-2.15/-9.1**
MC-MD	**-5.04/33.3**	**-6.13/-159.3**	**-6.06/-92.4**
*BL=Baseline, MC=Multicondition, MD=Missing data			

5. DISCUSSION

In this study, we carried out a comparison of three noise robust approaches in a large vocabulary ASR task using real noisy speech recordings from car and public place recordings. The main observations were that all the three methods improve the recognition accuracy substantially in poor SNR conditions when compared to a baseline system trained on clean speech. However, the DPMC and missing data systems give the best performance on high SNR conditions but the multicondition trained system outperforms the other systems on low SNR conditions where DPMC system performs the second best and missing data system the third best. In this study, multicondition training and the active model compensation method (DPMC) are evaluated separately whereas an approach called Noise Adaptive Training (NAT) makes it possible to combine effectively multicondition training with a VTS based model compensation technique. NAT yielded results that were superior both over standard multicondition training and VTS in the Aurora 2 digit recognition task [10].

The challenge in the recognition of the far recordings is not only due to the relative increase in noise level but also the convolutive channel effect on the speech signal. This is caused by the different types of microphones used in the recordings and the increasing effect of reverberation relative to direct sound as the microphone distance is increased. The multicondition trained system performs considerably better on mid and far recordings because it has a clear advantage of learning the respective channel distortions during training. DPMC and missing data systems are, on the other hand, only capable of compensating the additive noise, not the convolutional channel distortion. Had any convolutional noise robust method also been used such as cepstral mean subtraction (CMS), the baseline, DPMC and missing data systems would have shown significant performance improvements on mid and far recordings. CMS would have also improved the multicondition system performance but not on the same scale as the other systems. However, the standard version of CMS does not cope with the parameter reversibility demand of DPMC, thus a modified CMS has been introduced to be used with DPMC [19]. We intend to test the modified CMS in the future studies.

In the present study, the multicondition training set contained clean and noisy data from the same SPEECON corpus that was used to construct the evaluation sets. This has given some advantage for the multicondition trained system over the more flexible PMC and missing data approaches relying on adaptive compensation. While the settings in the collection of SPEECON are realistic and represent everyday use of speech technology in common environments with commonly used microphone settings, they are, however, standardized and do not represent all the variables that may be encountered in real use of ASR devices. It is noteworthy that due to the restricted corpus, we did not have noise types with considerable mismatch to the multicondition trained models.

The performance differences between the public place and car noises on each system is noteworthy. The car noise is more difficult for the baseline system to recognize which is also reflected to the performance of noise robust methods. This can be explained through the SNR differences since the car noise has significantly lower SNR on each microphone distance. Due to the limited frequency response of the far microphone on car noise recordings (see Section 3 for details), the respective SNR is effectively lower than measured.

But why is the multicondition system performance worse than the baseline on public place noise but better on car noise? Even though the training sets of multicondition and baseline systems are approximately the same, the multicondition training set includes considerably less clean data. The close recordings on car noise have such a low SNR which already affects considerably the clean trained baseline performance. Therefore the multicondition system has a performance advantage over the baseline. However, the close recordings on public place noise have higher SNR which, based on our experience, corresponds almost to the clean training condition and does not considerably degrade the baseline recognizer performance; rather the baseline recognizer has an advantage of having more clean data in the training set that matches well to the public place close recordings.

Missing data reconstruction performance was lower than the DPMC performance in most test conditions, but the differences were larger in public than in car environments. This is likely because the reconstruction performance is sensitive to mask estimation errors as discussed in [14]. The mask estimation method used in this work assumes the noise interference is stationary, which approximately holds for the car engine noise, but does not hold in public environments where the noise is often time-varying (e.g. speech, music, clatter). It is therefore reasonable to assume the missing data reconstruction performance would improve with better noise mask estimation methods. Moreover, methods based on e.g. perceptual criteria [1] or using a Bayes classifier [17] would remove the need for noise estimation in missing data methods altogether.

6. ACKNOWLEDGEMENTS

The work was supported by the Academy of Finland in projects numbered 129674, 114369, 135003 and 123431, and by the Helsinki Graduate School in Computer Science and Engineering.

REFERENCES

[1] J. Barker, "Robust automatic speech recognition," in D. L. Wang and G. J. Brown, editors, *Computational auditory scene analysis*, Wiley-Interscience, 2006.

[2] A. S. Bregman, *Auditory scene analysis*. Address: MIT Press, 1990.

[3] M. Cooke, "A glimpsing model of speech perception in noise," in *JASA*, vol. 119, pp. 1562–1573, 2006.

[4] M. P. Cooke, P. D. Green, L. Josifovski, and A. Vizinho, "Robust automatic speech recognition with missing and unreliable acoustic data," in *Speech Comm.*, vol. 34, 267–285, 2001.

[5] R. C. van Dalen, and M. J. Gales, "Covariance modelling for noise-robust speech recognition," in *Interspeech*, 2008.

[6] M. J. Gales, "Semi-tied covariance matrices for hidden Markov models," in *IEEE Trans. SAP*, vol. 7 pp. 272–281, 1999.

[7] M. J. Gales, *Model-based techniques for noise robust speech recognition*. Ph.D thesis, Univ. of Cambridge, 1995.

[8] T. Hirsimäki, M. Creutz, V. Siivola, M. Kurimo, S. Virpioja, and J. Pylkkönen, "Unlimited vocabulary speech recognition with morph language models applied to Finnish," in *Computer Speech and Language*, vol. 20, pp. 515–541, 2006.

[9] D. Iskra, B. Grosskopf, K. Marasek, H. van den Heuvel, and A. Kiessling, "SPEECON - speech databases for consumer devices: Database specification and validation," in *Proc. LREC*, pp. 329–333, 2002.

[10] O. Kalinli, M. Seltzer, and A. Acero. "Noise adaptive training using a vector Taylor series approach for noise robust automatic speech recognition," in *ICASSP*, pp. 3825–3828, 2009.

[11] P. J. Moreno, B. Raj, and R. M. Stern, "A Vector Taylor Series approach for environment-independent speech recognition," in *ICASSP*, vol. 2, pp. 733–736, 1996.

[12] R. W. Morris, and M. E. Deisher, "Efficient second-order adaptation for large vocabulary distributed speech recognition," in *ICASSP*, vol. 1, pp. 205–208, 2002.

[13] J. Pylkkönen, and M. Kurimo, "Duration modeling techniques for continuous speech recognition," in *Interspeech*, pp. 385–388, 2004.

[14] B. Raj and R. M. Stern, "Missing-feature approaches in speech recognition," in *IEEE Signal Processing Magazine*, vol. 22, pp. 101–116, 2005.

[15] B. Raj, M. L. Seltzer, and R. M. Stern, "Reconstruction of missing features for robust speech recognition," in *Speech Comm.*, vol. 43, pp. 275–296, 2004.

[16] U. Remes, K. J. Palomäki, and M. Kurimo, "Missing feature reconstruction and acoustic model adaptation combined for large vocabulary continuous speech recognition," in *EUSIPCO*, 2008.

[17] M. L. Selzer, B. Raj, and R. M. Stern, "Bayesian classifier for spectrographic mask estimation for missing feature speech recognition," in *Speech Comm.*, vol. 43, pp. 379–393, 2004.

[18] V. Siivola, and B. Pellom, "Growing an n-gram language model," in *Interspeech*, pp. 1309–1312, 2005.

[19] H. Yamamoto, T. Kosaka, M. Yamada, Y. Komori, and M. Fujita, "Fast speech recognition algorithm under noisy environment using modified CMS-PMC and improved IDMM+SQ," in *ICASSP*, vol. 2, pp. 847–850, 1997.