About Combining Forward and Backward-Based Decoders for Selecting Data for Unsupervised Training of Acoustic Models

Denis Jouvet, Dominique Fohr

About Combining Forward and Backward-Based Decoders for Selecting Data for Unsupervised Training of Acoustic Models

Denis Jouvet[1,2,3], *Dominique Fohr*[1,2,3]

Speech Group, LORIA
[1]Inria, Villers-lès-Nancy, F-54600, France
[2]Université de Lorraine, LORIA, UMR 7503, Villers-lès-Nancy, F-54600, France
[3]CNRS, LORIA, UMR 7503, Villers-lès-Nancy, F-54600, France
{denis.jouvet, dominique.fohr} @loria.fr

Abstract

This paper introduces the combination of speech decoders for selecting automatically transcribed speech data for unsupervised training or adaptation of acoustic models. Here, the combination relies on the use of a forward-based and a backward-based decoder. Best performance is achieved when selecting automatically transcribed data (speech segments) that have the same word hypotheses when processed by the Sphinx forward-based and the Julius backward-based transcription systems, and this selection process outperforms confidence measure based selection. Results are reported and discussed for adaptation and for full training from scratch, using data resulting from various selection processes, whether alone or in addition to the baseline manually transcribed data. Overall, selecting automatically transcribed speech segments that have the same word hypotheses when processed by the Sphinx forward-based and Julius backward-based recognizers, and adding this automatically transcribed and selected data to the manually transcribed data leads to significant word error rate reductions on the ESTER2 data when compared to the baseline system trained only on manually transcribed speech data.

Index Terms: Unsupervised training, combining recognizer outputs, data selection, LVCSR, speech recognition

1. Introduction

Acoustic speech models are one of the key components of a speech recognition system, along with the pronunciation lexicons and the language models. However, large amount of transcribed speech data are necessary for building good acoustic speech models. Such databases require a lot of manpower for manually transcribing the speech material; this is time consuming and expensive.

Consequently several studies have been devoted to investigating approaches to avoid or to limit the requirements on manually transcribed data for training acoustic models. The basic idea of such approaches is to automatically transcribe available speech material, and then possibly select the most reliably transcribed segments as a new training set, or as an extension to an available manually transcribed training set. For example in [1], lattice-based confidence measures are used to select the transcribed data to be used. As language models play an important role for speech transcription, a particular attention was paid on them in [2]. Along a similar path, lightly supervised and unsupervised acoustic model training was investigated in [3] and [4]. When available, close captions provide a useful hint for selecting automatically transcribed data (as for example selecting segments where automatically transcribed data matches with close caption). Lightly supervised training data was also used in discriminative training for improving broadcast news transcription [5]. Some refinements of the selection process have been proposed in [6] by carrying data selection at the state level. Finally unsupervised training have been applied on large data sets, as for example on thousands of hours of Arabic data [7],[8]; and [9] has compared the behavior of unsupervised training with supervised training on exactly the same data.

In a previous study we have investigated the combination of forward-based and backward-based speech recognition systems for improving speech recognition performance [10]. A detailed analysis of the behavior of the systems has shown that when the forward-based and the backward-based decoders provide the same word hypotheses, these common word hypotheses are correct in more than 90% of the cases [11]. Hence the goal of this paper which is to investigate how such behavior can help for selecting data for unsupervised training of acoustic models.

The paper is organized as follows. Section 2 presents the speech corpora used. Section 3 details the speech transcription systems. Section 4 presents methods for selecting segments of automatically transcribed data. Then, Section 5 details the results achieved when such selected data are used for adapting or for training acoustic models.

2. Speech corpora

The speech corpora used in the experiments come from the ESTER2 [12] and the ETAPE [13] evaluation campaigns, and the EPAC [14],[15] project. The ESTER2 and EPAC data are French broadcast news collected from various radio channels, thus they contain prepared speech, plus interviews. A large part of the speech data is of studio quality, and some parts are of telephone quality. On the opposite, the ETAPE data corresponds to debates collected from various radio and TV channels, thus contains mainly spontaneous speech.

The acoustic models used later in the experiments are trained using ESTER2, ETAPE and EPAC speech data. The baseline models are trained using the manually transcribed speech data from the ESTER2 and ETAPE training sets, as well as the manually transcribed data from the EPAC corpus; this amounts to almost 300 hours of signal and almost 4 million running words. Other models are built, from scratch or through adaptation of the baseline models, to investigate the impact of various selection processes for introducing automatically transcribed data into the unsupervised adaptation/training process. The non-transcribed part of the EPAC corpus, which amounts for about 1377 hours of signal is used for this purpose.

The development and test sets of the ESTER2 data are used for performance evaluation in the experiments reported below. The results (word error rates) are given for the non-African radios of the ESTER2 development set (about 42,000 running words), and for the non-African radios of the ESTER2 test set (about 63,000 running words). As the EPAC data does not contain African radios, there is no motivation in evaluating the performance in mismatch conditions; hence we have focused the evaluations on non-African radios. Performance evaluation on the ESTER2 data was carried on using the sclite tool [16] according to the ESTER2 campaign protocol.

3. Speech transcription systems

The speech recognition systems used in the following experiments are part of the set of forward-based and backward-based decoding systems that were studied and combined for improved speech transcription [10].

The speech transcription systems used in the experiments rely on a common diarization step and, on the one hand, on the Sphinx toolkit [17], and on the other hand, on the HTK toolkit [18] and the Julius decoder [19]. The diarization step associates to each speech segment, information about automatically identified speech quality, speaker gender and speaker identity label. Identified speech quality and speaker gender are used in the unsupervised gender adaptation of the studio and telephone quality acoustic models.

3.1. Forward Sphinx-based transcription system

The Sphinx-based transcription system relies on a lexicon of about 95,000 words and a trigram language model, The pronunciation lexicons were obtained using the pronunciation variants present in the BDLEX [20] lexicon and in in-house pronunciation lexicons; then, for the remaining words, the pronunciation variants were obtained automatically using both JMM-based and CRF-based Grapheme-to-Phoneme converters [21]. The trigram language model was trained using the SRILM tools [22] and various text corpora (more than 1,500 million words from newspapers, French Gigaword corpus [23], web data and manual transcriptions of radio broadcast shows).

The acoustic models are specific to gender (male vs. female) and speech quality (studio vs. telephone). HTK [18] MFCC features are used, plus their first and second order temporal derivatives, yielding 39 coefficient input vectors. Context-dependent phoneme units are used, and the baseline system has a total of 7,500 shared densities (senones), each of them having 64 Gaussian components. The first decoding pass does a decoding of each audio segment using the most adequate acoustic model (according to estimated speech quality and gender). The second decoding pass takes benefit from unsupervised VTLN adaptation of the features and MLLR adaptation of the acoustic models.

This system is also used as baseline, and its performance is reported in Table 2 (*Baseline*).

3.2. Backward Sphinx-based transcription system

A similar system, but based on a reverse processing approach, has also been developed: the frames of each audio segment are given to the training tool and to the decoder in a reverse time order (i.e. last frame of each audio segment is given first). The pronunciation of each word in the lexicon is also reversed, and language models are re-estimated after reversing the sequences of words of all the text sentences. The corresponding reverse (backward-based) system achieves similar performance as the standard (forward-based) system, however, these two Sphinx-based systems (forward vs. backward) do not make the same recognition errors.

3.3. Backward Julius-based transcription system

The Julius (backward-based) decoder uses acoustic models dependent on the speech quality (studio vs. telephone). Context-dependent phoneme units are used, and are modeled with 6,000 shared states/densities, and each mixture density has 62 Gaussian components. This transcription system runs also in two passes; and the second transcription pass takes benefit from SAT (Speaker Adaptive Training) adapted models.

The Julius decoder relies on a forward-backward process. A forward pass uses a bigram and generates a word graph; then, a backward A* pass explores this graph guided by a reverse 4-gram language model.

HTK MFCC features are used, and an HLDA transform is applied on windows of 9 acoustic feature vectors to provide the 40 input modeling coefficients. The phoneme units chosen ignore the aperture of the vowels (for example, open /ɛ/ and close /e/ are merged in a same unit).

4. Selection of unsupervised transcripts

In a previous study it was observed that when the forward-based and the backward-based decoders provide a common word hypothesis, this word hypothesis is correct in more than 90% of the cases [11]. Hence, the goal of this study which investigates how to take benefit of such a behavior for unsupervised training.

The speech transcription systems described above were applied on the non-transcribed part of the EPAC corpus (about 1377 hours of signal). Then automatically transcribed segments were selected for adapting or for training acoustic models. Two constrains were applied in the selection process: each selected segment must either correspond to more than 10 words or be preceded and followed by more than 300 ms of non-speech data. Non-speech data is identified by silence or filler units. Moreover, whenever possible, some non-speech data is kept before and after the selected speech segment (this was set mandatory for short segments, and optional for long segments).

The main criterion we want to investigate is the impact of selecting speech segments that correspond to the same recognition hypotheses when decoded by a forward-based and a backward-based system. Two cases are considered: selecting segments that correspond to the same word hypotheses with the Sphinx forward and the Sphinx backward systems, and selecting segments that correspond to the same word hypotheses with the Sphinx forward and the Julius systems. Note that the speech segments are not determined a priori, but are defined from the common word hypothesis sequences that result from the comparison of the recognizer outputs.

These proposed selection processes are compared to the usage of confidence measure. For this purpose the computation of word posterior probabilities was implemented in the Julius decoder. The word posterior probabilities [24] are computed from the word graph using a forward-backward based method, and are used as confidence measures. In the reported experiments, words are selected only if their

confidence measure is above a given threshold (two threshold values are considered: 0.6 and 0.4). Before applying this threshold-based selection process, a light smoothing process is applied on the computed word confidence measures to avoid rejecting a low confidence word occurring between two high confidence words.

Table 1. *Amount of speech data (in hours of signal) in manually transcribed data and in automatically transcribed data after selection according to the various methods.*

Transcription / Selection method	Manual	Automatic		
	All	Male	Fem.	All
(*a*) *Baseline*	*300*	--	--	--
(*b*) Sphinx forward & backward	--	539	160	699
(*c*) **Sphinx & Julius**	**--**	**420**	**122**	**542**
(*d*) Sphinx & Julius & cm >= 0.6	--	209	61	270
(*e*) Julius only & cm >= 0.6	--	303	86	389
(*f*) Sphinx & Julius & cm >= 0.4	--	288	84	372
(*g*) Julius only & cm >= 0.4	--	456	129	585

Table 1 reports the amount of speech data selected in automatically transcribed data by the various selection methods. It clearly appears that selecting segments having a common decoding with the Sphinx-based and Julius-based decoders (*line c*) leads to a much larger amount of selected data than using Julius only and applying a reasonable threshold on the confidence measures (e.g. 0.6 – *line e*). To reach a similar amount of selected data, the selection threshold has to be lowered significantly (down to 0.4 – *line g*). The amount of manually transcribed data available in the baseline training data is also reported.

5. Using unsupervised transcribed data in acoustic model training

Unsupervised training experiments have been conducted using subsets of automatically transcribed speech data; the subsets were obtained according to the various selection procedures described in Section 4. Several aspects are considered, whether the selected data are used alone, or in addition to the baseline training set (which was manually transcribed); and whether the data are used for adapting the acoustic models or for re-doing a full training of the models from scratch.

5.1. Adding automatically transcribed data to the manually transcribed baseline training set

In the first set of experiments, the selected data is added to the baseline manually transcribed training set for the gender adaptation (MLLR + MAP) of the generic models (one for studio quality speech, and one for telephone quality model). For the *Baseline* models only the manually transcribed initial training set is used for gender adaptation.

Table 2, as well as following tables, reports the total amount of data used for model adaptation, or model full training, depending on experiments. This corresponds to the amount of automatically transcribed data selected, plus possibly the manually transcribed data used for the baseline model (in case of extended training sets).

Table 2 shows that selecting automatically transcribed speech data segments that correspond to common word hypotheses by the Sphinx forward-based and the Julius backward-based speech transcription system leads to significant error rate reduction after gender adaptation on the extended training set (baseline training set plus selected data – *line c*), compared to the baseline system (where gender adaptation is carried on using the baseline training set only – *line a*).

Table 2. *Word error rates on ESTER2 Dev and Test sets after adaptation using the extended training sets resulting from various selection methods.*

Selection method	Adapt. data	ESTER2 Dev	ESTER2 Test
(*a*) *Baseline*	*300 h*	*20.73%*	*21.17%*
(*b*) Sphinx forward & backward	999 h	20.48%	20.90%
(*c*) **Sphinx & Julius**	**842 h**	**20.29%**	**20.83%**
(*d*) Sphinx & Julius & cm >= 0.6	570 h	20.73%	21.15%
(*e*) Julius only & cm >= 0.6	689 h	20.48%	20.93%
(*f*) Sphinx & Julius & cm >= 0.4	672 h	20.74%	20.95%
(*g*) Julius only & cm >= 0.4	885 h	20.62%	20.91%

Selecting words that correspond to a common decoding and also have a confidence measure above a reasonable threshold (for example cm>=0.6), reduces the amount of selected data; however, the selected subset may possibly become too similar to the baseline training data. The combined selection criterion (common decoding plus high enough confidence measure) does not provide any improvement over the baseline (threshold 0.6 – *line d*) or just a small improvement (on the test set) when a smaller threshold is used (threshold 0.4 – *line f*).

An intermediate improvement is achieved when the data is selected using word hypotheses common to the Sphinx forward-based and backward-based systems (*line b*), or only according to the confidence measure (*lines e & g*). Although more data are selected using Julius only and low confidence measure threshold (*line g*) the results are not as good as those provided through a selection relying on a common decoding with Sphinx and Julius (*line c*).

A detailed analysis of the errors was conducted to determine the amount of errors specific to each system or common to two systems, in order to apply the McNemar test to compare the systems two by two [25]. The McNemar test showed that many differences are significant. For example, relying on common decoded segments with the Sphinx and Julius systems (*line c*) leads to results significantly better than the baseline (*line a* - p-value=0.0006 on Dev and p-value=0.003 on Test), and, on the Dev set, the result is also better than using only a confidence-based criteria leading to the same amount of data (*line g* – p-value=0.003 on Dev). Moreover, using the common decoding avoids having to choose which threshold on the confidence measure is the best.

Table 3 shows that when a full training of the acoustic models from scratch is carried out using the various extended data sets, the achieved results are not as good as those obtained before (where the data was used only for adaptation). Consequently there is no benefit in retraining the base model from scratch using the extending data sets.

Table 3. *Word error rates on ESTER2 Dev and Test sets after full training of acoustic models using the extended training sets resulting from various selection methods*.

Selection method	Train/ adapt.	ESTER2 Dev	ESTER2 Test
(*a*) *Baseline*	*300 h*	*20.73%*	*21.17%*
(*c*) **Sphinx & Julius**	**842 h**	20.59%	21.11%
(*d*) Sphinx & Julius & cm >= 0.6	570 h	20.85%	21.06%
(*e*) Julius only & cm >= 0.6	689 h	20.77%	21.23%
(*f*) Sphinx & Julius & cm >= 0.4	672 h	20.79%	20.90%
(*g*) Julius only & cm >= 0.4	885 h	20.49%	20.93%

5.2. Adaptation vs. training of acoustic models

Here, gender adaptation of the studio and telephone generic models and full training from scratch of the acoustic models are compared for different model sizes. Evaluations are carried out using the extended training set obtained by adding speech segments having a common decoding with the Sphinx and Julius systems (corresponding to *lines c* in previous tables).

Table 4. *Word error rates on ESTER2 Dev and Test sets for various model sizes after gender adaptation only or full training of acoustic models using the extended training set (obtained by adding speech segments having common word hypotheses with Sphinx and Julius).*

A) 7500 senones / 64 gauss.	Train/ adapt.	ESTER2 Dev	ESTER2 Test
Baseline (⇔ line a, Tables 2 & 3)	*300 h*	*20.73%*	*21.17%*
Adaptation (⇔ *line c, Table 2*)	**842 h**	**20.29%**	**20.83%**
Full Training (⇔ *line c, Table 3*)	842 h	20.59%	21.11%

B) 7500 senones / 128 gauss.	Train/ adapt	ESTER2 Dev	ESTER2 Test
Baseline	*300 h*	*20.87%*	*21.32%*
Adaptation	**842 h**	**20.47%**	**20.91%**
Full Training	842 h	20.41%	20.87%

C) 15000 senones / 64 gauss.	Train/ adapt	ESTER2 Dev	ESTER2 Test
Baseline	*300 h*	*21.56%*	*22.05%*
Adaptation	**842 h**	**20.56%**	**21.23%**
Full Training	842 h	20.92%	21.22%

Results reported in Table 4 show that the baseline training set is not large enough to handle a large increase in the amount of parameters of the acoustic models. The baseline results degrade when either the number of senones (shared densities) or the number of components per density is doubled. Using the extended training set (after adding automatically transcribed data corresponding to Sphinx and Julius common word hypotheses) improves significantly the performance for each model. However, after adaptation on the extended training set, the larger acoustic models do not outperform the initial baseline model (7500 senones / 64 gauss.). Globally, full training from scratch using the extended training sets does not provide significantly better results than just adaptation with these extended training sets.

5.3. Training from automatically transcribed data only

This set of experiments aims at analyzing the quality of the selected automatically transcribed data when used alone for training acoustic models.

Table 5 shows the results obtained when selecting speech segments according to various selection processes described before. The important point to note is that although the acoustic models are trained from scratch using only automatically transcribed data, the word error rates are only about 1% worse than that achieved with a training relying on a large manually transcribed training set (*cf. Baseline in previous tables*).

Table 5. *Word error rates on ESTER2 Dev and Test sets after full training of acoustic models using only automatically transcribed data*.

Selection method	Train/ adapt	ESTER2 Dev	ESTER2 Test
(*c*) **Sphinx & Julius**	**542 h**	**21.56%**	**22.49%**
(*d*) Sphinx & Julius & cm >= 0.6	270 h	22.81%	23.57%
(*e*) Julius only & cm >= 0.6	389 h	22.26%	23.00%
(*f*) Sphinx & Julius & cm >= 0.4	372 h	22.27%	23.47%
(*g*) Julius only & cm >= 0.4	585 h	21.68%	22.14%

6. Conclusion

This paper has presented and analyzed the usage of various decoders (typically a Sphinx forward-based system and a Julius backward-based system) for optimizing the selection of automatically transcribed data for adapting acoustic models. Selecting automatically transcribed data that have common word hypotheses with the Sphinx and Julius based decoders leads to better results than when selecting automatically transcribed data according to a confidence measure criterion.

The best speech recognition performance is achieved when the extended training set (i.e. manually transcribed training set plus addition of automatically transcribed data) is used for gender adaptation of the studio and telephone quality generic models. A significant reduction in the word error rate (about 0.4% absolute, Mc Nemar p-value < 0.005) is achieved on the ESTER2 Dev and Test sets with respect to the baseline model which was trained using about 300 hours of manually transcribed speech.

Moreover, full training from scratch using only automatically transcribed data, leads to results which are rather close to the baseline results. This makes possible the application of automatic transcription on new types of data; for example transcription of large bandwidth data (with telephone-based models, after signal filtering) in view of later training large bandwidth models. This should help taking benefit of the large telephone speech corpora that have been recorded and manually transcribed in the last decades in many languages.

7. References

[1] Kemp, T. and Waibel, A., "Unsupervised training of a speech recognizer: recent experiments", in *Proc. EUROSPEECH'99*, 6th *European Conference on Speech Communication and Technology*, Budapest, Hungary, pp. 2725-2728, September 1999.

[2] Lamel, L., Gauvain, J.-L. and Adda, G., "Unsupervised acoustic model training", in *Proc. ICASSP'2002, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Orlando, Florida, USA, pp. 877-880, May 2002.

[3] Nugyen, L. and Xiang, B., "Light supervision in acoustic training", in *Proc. ICASSP'2004, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Montreal, Quebec, Canada, pp. 185-188, May 2004.

[4] Lamel, L., Gauvain, J.-L. and Adda, G., "Lightly supervised and unsupervised acoustic model training", *Computer Speech and Language*, 16(1):115-229, 2002.

[5] Chan, H.Y. and Woodland, P., "Improving broadcast news transcription by lightly supervised discriminative training", in *Proc. ICASSP'2004, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Montreal, Quebec, Canada, pp. 737-740, May 2004.

[6] Gollan, C., Hahn, S., Schlüter, R. and Ney, H., "An improved method for unsupervised training of LVCSR systems", in *Proc. INTERSPEECH 2007, 8th Annual Conf. of the Int. Speech Communication Association*, Antwerp, Belgium, pp. 2101-2104, August 2007.

[7] Ma, J., Matsoukas, S., Kimball, O. and Schwartz, R., "Unsupervised training on large amounts of broadcast news data," in *Proc. ICASSP'2006, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Toulouse, France, vol. 3, pp. 1057-1060, May 2006.

[8] Ma, J. and Matsoukas, S., "Unsupervised training on a large amount of Arabic broadcast news data", in *Proc. ICASSP'2007, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Honolulu, Hawaii, USA, April 2007.

[9] Ma, J. and Schwartz, R., "Unsupervised versus supervised training of acoustic models", in *Proc. INTERSPEECH'2008, 9th Annual Conf. of the Int. Speech Communication Association*, Brisbane, Australia, pp. 2374-2377, September 2008.

[10] Jouvet, D. and Fohr, D., "Combining forward-based and backward-based decoders for improved speech recognition performance"; in *Proc. INTERSPEECH'2013, 14th Annual Conf. of the Int. Speech Communication Association*, Lyon, France, August 2013.

[11] Jouvet, D. and Fohr, D., "Analysis and combination of forward and backward based decoders for improved speech transcription"; in *Proc. TSD'2013, 16th Int. Conf. on Text, Speech and Dialogue*, Pilsen, Czech Republik, September 2013.

[12] Galliano, S., Gravier, G. and Chaubard, L., "The Ester 2 evaluation campaign for rich transcription of French broadcasts", in *Proc. INTERSPEECH'2009, 10th Annual Conf. of the Int. Speech Communication Association*, Brighton, UK, pp. 2583-2586, September 2009.

[13] Gravier, G., Adda, G., Paulsson, N., Carré, M., Giraudel, A. and Galibert, O., "The ETAPE corpus for the evaluation of speech-based TV content processing in the French language", in *Proc. LREC'2012, Int. Conf. on Language Resources, Evaluation and Corpora*, Istanbul, Turkey, May 2012.

[14] Estève, Y., Bazillon, T., Antoine, J.-Y., Béchet, F. and Farinas, J., "The EPAC corpus: Manual and automatic annotations of conversational speech in French broadcast news", in *Proc. LREC'2010, European Conf. on Language Resources and Evaluation*, Valetta, Malta, May 2010.

[15] Corpus EPAC: Transcriptions orthographiques, catalogue ELRA (http://catalog.elra.info), reference ELRA-S0305.

[16] NIST evaluation tools: http://www.itl.nist.gov/iad/mig//tools/

[17] Sphinx. [Online]: http://cmusphinx.sourceforge.net/, 2011.

[18] HTK. [Online]: http://htk.eng.cam.ac.uk/

[19] Julius. [Online]: http://julius.sourceforge.jp/en_index.php

[20] de Calmès, M. and Pérennou, G., "BDLEX : a Lexicon for Spoken and Written French." in *Proc. LREC'1998, 1st Int. Conf. on Language Resources & Evaluation*, pp.1129-1136, Grenada, Spain, May 1998.

[21] Illina, I., Fohr, D. and Jouvet, D., "Grapheme-to-Phoneme Conversion using Conditional Random Fields", in *Proc. INTERSPEECH'2011, 12th Annual Conf. of the Int. Speech Communication Association*, Florence, Italy, August 2011.

[22] Stolcke, A., "SRILM - An Extensible Language Modeling Toolkit", in *Proc. ICSLP'2002, Int. Conf. on Spoken Language Processing*, Denver, Colorado, September 2002.

[23] Mendonça, A., Graff, D. and DiPersio, D., "French Gigaword Second Edition", Linguistic Data Consortium, Philadelphia, 2009.

[24] Wessel, F., Schlüter, R., Macherey, K. and Ney, H., "Confidence measures for large vocabulary continuous speech recognition," *IEEE Trans. SAP, Speech and Audio Processing*, vol. 9, pp. 288–298, 2001

[25] Gillick, L. and Cox, S. J., "Some statistical issues in the comparison of speech recognition algorithms"; in *Proc. ICASSP89, Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 532-535, 1989.