AUTOMATIC VS. MANUAL TOPIC SEGMENTATION AND INDEXATION IN BROADCAST NEWS

R. Amaral$^{(1,2,3)}$, *H. Meinedo*$^{(1,3)}$, *D. Caseiro*$^{(1,3)}$, *I. Trancoso*$^{(1,3)}$, *J. Neto*$^{(1,3)}$ *

(1) Instituto Superior Técnico
(2) Instituto Politécnico de Setúbal
(3) L^2F - Spoken Language Systems Lab, INESC-ID
{ramaral,meinedo,dcaseiro,imt,jpn}@l2f.inesc-id.pt
http://www.l2f.inesc-id.pt

ABSTRACT

This paper describes the latest progress in our work on Broadcast News for European Portuguese. The central modules of our media watch system that matches the topic of each news story against the user preferences registered in the system are: audio pre-processing, speech recognition and topic segmentation and indexation. The main focus of the paper is on the impact of the errors made by the earlier modules in the last ones. This impact is in our opinion an essential diagnostic tool for the improvement of the overall pipeline system.

1. INTRODUCTION

In a media watch system dealing with selective dissemination of Broadcast News, there are several components whose performance influences the next stages. In the system developed at our lab, the first of these components is the audio pre-processing (APP) module which performs speech/non-speech classification, speaker segmentation, speaker clustering, and gender and background conditions classification. The second component is the automatic speech recognition (ASR) module that converts the segments classified as speech into text. The XML file produced by these earlier modules includes not only the transcribed text, but also additional information such as the segment duration, the acoustic background classification (e.g. clean/music/noise), the speaker gender and the identification of the speaker cluster. The third component is the topic segmentation (TS) module that takes as input this XML file and splits the broadcast news program show into constituent stories. Last in this pipeline of modules is the topic indexation (TI) module that assigns one or multiple topics selected from a thematic thesaurus, thus adding more metadata to the original XML file. The goal of this paper is to study the influence of the performance of the earlier modules on the next ones.

The use of a thematic thesaurus for indexation was introduced by RTP, the Portuguese public broadcast company, and our partner in the past European project ALERT. This thesaurus follows rules which are generally adopted within EBU (European Broadcast Union) and has been used by RTP since 2002 in its daily manual indexation task. It has a hierarchical structure that covers all possible topics, with 22 thematic areas in the first level, and up to 9 lower levels. In our system, we implemented only 3 levels, which are enough to represent the user profile information that we need to match against the topics produced by the indexation module.

This paper is structured into four main sections, each one devoted to one of the four modules. Rather than lumping all the results together, we will present them individually for each section, in order to be able to better compare the oracle performance of each module with the one in which all previous components are automatic. Before describing each module and the corresponding results, we shall describe the corpus that served as the basis for this study. The last section summarizes the main conclusions and future work.

2. THE EUROPEAN PORTUGUESE BN CORPUS

The European Portuguese Broadcast News corpus, collected in close cooperation with RTP, involves different types of news shows, national and regional, from morning to late evening, including both normal broadcasts and specific ones dedicated to sports and financial news. The corpus is divided into 3 main subsets:

- SR (Speech Recognition) - The SR corpus contains around 61h of manually transcribed news, collected during a period of 3 months, with the primary goal of training acoustic models and adapting the language models of our ASR module. The corpus is subdivided into training (51h), development (6h), and evaluation sets (4h).

* H. Meinedo was sponsored by a FCT scholarship (SFRH/BD/6125/2001). This work was partially funded by FCT projects POSI/PLP/47175/2002 and POSC/PLP/58697/2004. INESC-ID Lisboa had support from the POSI program of the ”Quadro Comunitário de Apoio III”. The order of the first two authors was randomly selected.

- TD (Topic Detection) - The TD corpus contains around 300h of topic labeled news, collected during the following 9 months. All the data was manually segmented into stories or fillers (short segments spoken by the anchor announcing important news that will be reported later), and each story was manually indexed according to the thematic thesaurus. The corresponding orthographic transcriptions were automatically generated by our ASR module.

- JE (Joint Evaluation) - The JE corpus contains around 13h, corresponding to the last two weeks of the collection period. It was fully manually transcribed, both in terms of orthographic and topic labels. All the evaluation work described in this paper concerns this corpus. 17% was manually classified as being in the F0 focus condition (planned speech, no background noise, high bandwidth channel, native speech), whereas 61% is in the F4 condition (speech under degraded acoustical conditions).

3. AUDIO PRE-PROCESSING

The APP module includes five separate components: three for classification (speech/non-speech, gender and background), one for speaker clustering and one for acoustic change detection. All components are model-based, making extensive used of feed-forward fully connected Multi-Layer Perceptrons (MLP) trained with the back-propagation algorithm on the SR training corpus [1].

The speech/non-speech module is responsible for identifying audio portions that contain clean speech, and audio portions that instead contain noisy speech or any other sound or noise, such as music, traffic, etc. This serves two purposes: first, no time will be wasted trying to recognize audio portions that do not contain speech; second, it reduces the probability of speaker clustering mistakes.

Gender classification distinguishes between male and female speakers and is used to improve speaker clustering. By clustering separately each gender class we have a smaller distance matrix when evaluating cluster distances which effectively reduces the search space. It also avoids short segments having opposite gender tags being erroneously clustered together. Background status classification indicates if the background is clean, has noise or music. Although it could be used to switch between tuned acoustic models trained separately for each background condition, it is only being used for topic segmentation purposes. All three classifiers share the same architecture: an MLP with 9 input context frames, a hidden layer with 300 sigmoidal units, and an output unit which can be viewed as giving a probabilistic estimate of the input frame being speech or non-speech. When the acoustic change detector hypothesizes the start of a new segment, the first 300 frames of that segment are used to calculate the speech/non-speech, gender and background classification. Each classifier computes the decision with the highest average probability over all the frames.

The main goal of the acoustic change detector is to detect audio locations where speakers or background conditions have changed. It uses a hybrid two-stage algorithm. The first stage generates a large set of candidate change points which in the second stage are evaluated to eliminate the ones that do not correspond to true speaker change boundaries [1]. The first stage uses two complementary algorithms. It evaluates in the cepstral domain the similarity between two contiguous windows of fixed length using the symmetric Kullback-Liebler distance [2]. This is followed by an energy-based algorithm that detects when the median drops bellow the long term average. The second stage uses an MLP classifier, with a large 300-frame input context of acoustic features (12^{th} order PLP plus log energy) and a hidden layer with 150 sigmoidal units.

The goal of speaker clustering is to identify and group together all speech segments that were uttered by the same speaker. After the acoustic change detector signals the existence of a new boundary and the classification modules determine that the new segment contains speech, the first 300 frames of the segment are compared with all the clusters found so far, for the same gender. The segment is merged with the cluster with the lowest distance, provided it falls bellow a predefined threshold. The distance is also computed by one of the two gender-specific MLP classifiers.

3.1. Audio Pre-Processing Results

Table 1 summarizes the results for the components of the APP module computed over the JE corpus. Speech/non-speech, gender and background classification results are reported in terms of percentage of correctly classified frames for each class and accuracy, defined as the ratio between the number of correctly classified frames and the total number of frames. In order to evaluate the clustering, a bi-directional one-to-one mapping of reference speakers to clusters was computed (NIST rich text transcription evaluation script). The Q-measure is defined as the geometrical mean of the percentage of cluster frames belonging to the correct speaker and the percentage of speaker frames labeled with the correct cluster. Another performance measure is the DER (Diarization Error Rate) which is computed as the percentage of frames with an incorrect cluster-speaker correspondence.

Generally, these APP results are quite good, except in terms of background classification, which is a rather difficult task, and DER, for which results around 10% have been reported, obtained with state of the art speaker identification techniques like feature warping and model adaptations [3]. We are currently working on reducing the number of clusters belonging to the same speaker.

Our system was compared against the best algorithms evaluated in [4] applied to a common database, having achieved similar results in most categories.

Speech/		Speech	Non-speech	Accuracy
Non-Speech		98.9	54.9	95.4
Gender		Male	Female	Accuracy
		96.5	97.5	97.0
Background	Clean	Music	Noise	Accuracy
	77.4	79.0	71.4	73.5
Clustering		Q	Q map	DER
		76.9	84.7	29.8

Table 1. Audio Pre-Processing results.

4. AUTOMATIC SPEECH RECOGNITION

The second module in our pipeline system is a hybrid automatic speech recognizer [5] that combines the temporal modeling capabilities of Hidden Markov Models (HMMs) with the pattern discriminative classification capabilities of MLPs. The acoustic modeling combines phone probabilities generated by several MLPs trained on distinct feature sets: PLP (Perceptual Linear Prediction), Log-RASTA (log-RelAtive SpecTrAl) and MSG (Modulation SpectroGram). Each MLP classifier incorporates local acoustic context via an input window of 13 frames. The resulting network has two non-linear hidden layers with 1500 units each and 40 softmax output units (38 phones plus silence and breath noises). The vocabulary includes around 57k words. The lexicon includes multiple pronunciations, totaling 65k entries. The corresponding out-of-vocabulary (OOV) rate is 1.4%. The language model which is a 4-gram backoff model was created by interpolating a 4-gram newspaper text language model built from over 604M words with a 3-gram model based on the transcriptions of the SR training set with 532k words. The language models were smoothed using Knesser-Ney discounting and entropy pruning. The perplexity obtained in a development set is 112.9.

Our decoder is based on the Weighted Finite-State Transducer (WFST) approach to large vocabulary speech recognition [6]. In this approach, the search space is a large WFST that maps HMMs (or in some cases, observations) to words. This WFST is built by composing various components represented as WFSTs. In our case, the search space integrates the HMM/MLP topology transducer, the lexicon transducer and the language model one. Traditionally, this composition and subsequent optimization is done in an offline compilation step. A unique characteristic of our decoder is its ability to compose and optimize the various components of the system in runtime. A specialized WFST composition algorithm was developed [7] that composes and optimizes the lexicon and language model components in a single step. Furthermore, the algorithm can support lazy implementations so that only the fragment of the search space required in runtime is computed. This algorithm is able to perform true composition and determinization of the search space while approximating other operations such as pushing and minimization. This dynamic approach has several advantages relative to the static approach. The first one is memory efficiency, the specialized algorithm requires less memory than the explicit determinization algorithm used in the offline compilation step, moreover, since only a small fraction of the search space is computed, it also requires less runtime memory. This memory efficiency allows us to use large 4-gram language models in a single pass of the decoder. Other approaches are forced to use a smaller language model in the first pass and rescore with a larger language model. The second advantage is flexibility, the dynamic approach allows for quick runtime reconfiguration of the decoder since the original components are available in runtime and can be quickly adapted or replaced.

4.1. Confidence Measures

Associating confidence scores to the recognized text is essential for evaluating the impact of potential recognition errors. Hence, confidence scoring was recently integrated in the ASR module. In a first step, the decoder is used to generate the best word and phone sequence, including information about the word and phone boundaries, as well as search space statistics. Then, for each recognized phone, a set of confidence features are extracted from the utterance and from the statistics collected during decoding. The phone confidence features are combined into word level confidence features. Finally, a maximum entropy classifier is used to classify words as correct or incorrect. The word level confidence feature set includes various recognition scores (recognition score, acoustic score and word posterior probability [8]), search space statistics (number of competing hypotheses and number of competing phones), and phone log likelihood ratios between the hypothesized phone and the best competing one. All features are scaled to the $[0, 1]$ interval. The maximum entropy classifier [9] combines these features according to:

$$P(correct|w_i) = \frac{1}{Z(w_i)} exp[\sum_{i=i}^{F} \lambda_i f_i(w_i)] \qquad (1)$$

where w_i is the word, F is the number of features, $f_i(w_i)$ is a feature, $Z(w_i)$ is a normalization factor and λ_i are the model parameters. The detector was trained on the SR training corpus. When evaluated on the JE corpus, an equal-error-rate of 24% was obtained.

4.2. ASR Results with Manual and Automatic Pre-Processing

Table 2 presents the word error rate (WER) results on the JE corpus, for two different focus conditions (F0 and all conditions), and in two different experiments: according to the manual pre-processing (reference classifications and boundaries) and according to the automatic pre-processing defined by the APP module.

The performance is comparable in both experiments with only 0.7% absolute increase in WER. This increase

	% WER	
APP	F0	All
Manual	11.3	23.5
Automatic	11.6	24.2

Table 2. Speech recognition results.

can be explained by speech / non-speech classification errors, that is, word deletions caused by noisy speech segments tagged by APP as non-speech, and word insertions caused by noisy "silence" segments marked by APP as containing speech. The other source for errors is related to the differences between manual and automatic sentence-like unit (SU) boundaries. Since the APP tends to create larger than "real" SUs, the problem seems to be in the language model which is introducing erroneous words (mostly function words), connecting different SUs. A qualitative analysis indicates the following types of error:

- Errors due to severe vowel reduction. Vowel reduction, including quality change, devoicing and deletion, is specially important for European Portuguese, being one of the features that distinguishes it from Brazilian Portuguese and that makes it more difficult to learn for a foreign speaker. It may take the form of (1) intra-word vowel devoicing; (2) voicing assimilation; and (3) vowel and consonant deletion and coalescence. Both (2) and (3) may occur within and across word boundaries. Contractions are very common, with both partial or full syllable truncation and vowel coalescence. As a result of vowel deletion, rather complex consonant clusters can be formed across word boundaries. This type of error is strongly affected by high speech rate. The relatively high deletion rate may be partly attributed to severe vowel reduction and affects mostly (typically short) function words.

- Errors due to OOVs. This affects namely foreign names. It is known that one OOV term can lead to between 1.6 and 2 additional errors [10].

- Errors in inflected forms. This affects mostly verbal forms (Portuguese verbs typically have above 50 different forms, excluding clitics), and gender and number distinctions in names and adjectives. It is worth exploring the possibility of using some post-processing parsing step for detecting and hopefully correcting some of these agreement errors. Some of these errors are due to the fact that the correct inflected forms are not included in the lexicon.

- Errors around speech disfluencies. This is the type of error that is most specific of the spontaneous speech, a condition that is fairly frequent in the JE corpus. The frequency of repetitions, repairs, restarts and filled pauses is very high in these conditions, in agreement with values of one disfluency every 20 words cited in [11]. Unfortunately, the training corpus for Broadcast News included a very small representation of such examples.

- Errors due to inconsistent spelling of the manual transcriptions. The most common inconsistencies occur for foreign names and for entries spelled both as separate words and as a single word.

These WER results are worse than the ones that are quoted for other languages, such English (less than 16% with Real-Time performance [12]), a fact that can be partly attributed to the reduced amount of training data for European Portuguese.

5. TOPIC SEGMENTATION

The goal of TS module is to split the broadcast news show into the constituent stories. This may be done taking into account the characteristic structure of broadcast news shows [13]. They typically consist of a sequence of segments that can either be stories or fillers. The fact that all stories start with a segment spoken by the anchor, and are typically further developed by out-of-studio reports and/or interviews is the most important heuristic that can be exploited in this context. Hence, the simplest TS algorithm is the one that starts by defining potential story boundaries in every transition non-anchor / anchor. In the next step, the algorithm tries to eliminate stories that are too short, because of the difficulty of assigning a topic with so little transcribed material. In these cases, the short story segment is merged with the following one with the same speaker and background. Other heuristics are also adopted to avoid too long stories spoken only by the anchor, which may in fact include more than one story without further developments.

The identification of the anchor is done on the basis of the speaker clustering information, as the cluster with the largest number of turns. A minor refinement was recently introduced to account for the cases where there are two anchors (although not present in the JE corpus).

5.1. Topic Segmentation Results with Manual and Automatic Prior Processing

The evaluation of the topic segmentation was done using the standard measures Recall (% of detected boundaries), Precision (% of marks which are genuine boundaries) and F-measure (defined as $2RP/(R+P)$). Table 3 shows the TS results, using the Recall, Precision, and F-measure metrics, as well the metric adopted in the 2001 Topic Detection and Tracking benchmark NIST evaluation, with the same cost values of miss and false alarms [14]. These results, together with the field trials we have conducted [15], show that boundary deletion is a critical problem. In fact, our very simple TS algorithm has several pitfalls: it fails when all the story is spoken by the anchor, without further reports or interviews, leading to a merge with the next story; ii) it fails when the filler is not detected by a speaker / background condition change, also leading to a merge with the next story (19% of the program events are fillers); iii) it fails when there is a special anchor for a part of the broadcast (i.e. sports anchor), although in this case one could argue that all the stories are about the same generic topic; iv) it fails when the anchor(s) is not correctly identified.

APP	ASR	Recall	Precision	F-meas.	Cost
Manual	Manual	79.0	60.3	67.2	0.78
Manual	Auto	76.6	60.3	66.3	0.75
Auto	Auto	70.2	56.4	61.6	0.66

Table 3. Topic Segmentation results.

6. TOPIC INDEXATION

Topic identification is a two-stage process, that starts with the detection of the most probable top-level story topics and then finds for those topics all the second and third level descriptors that are relevant for the indexation.

For each of the 22 top-level domains, topic and non-topic unigram language models were created using the stories of the TD corpus which were pre-processed in order to remove function words and lemmatize the remaining ones. Topic detection is based on the log likelihood ratio between the topic likelihood $p(W/T_i)$ and the non-topic likelihood $p(W/\overline{T_i})$. The detection of any topic in a story occurs every time the correspondent score is higher than a predefined threshold. The threshold is different for each topic in order to account for the differences in the modeling quality of the topics.

In the second step, we count the number of occurrences of the words corresponding to the domain tree leafs and normalize these values with the number of words in the story text. Once the tree leaf occurrences are counted, we go up the tree accumulating in each node all the normalized occurrences from the nodes below [16]. The decision of whether a node concept is relevant for the story is made only at the second and third upper node levels, by comparing the accumulated occurrences with a predefined threshold.

6.1. Topic Indexation Results with Manual and Automatic Prior Processing

In order to conduct the topic indexation experiments we started by choosing the best threshold for the word confidence measure as well as for the topic confidence measure. The tuning of these thresholds was done with the development corpus in the following manner: the word confidence threshold was ranged from 0 do 1, and topic models were created using the correspondent topic material available. Obviously higher threshold values decrease the amount of automatic transcriptions available to train each topic. Topic indexation was then performed in the development corpus in order to find the topic thresholds corresponding to the best topic accuracy (91.9%). The use of these confidence measures led to rejecting 42.0% of the original topic training material.

Once the word and topic confidence thresholds were defined, the evaluation of the indexation performance was done for all the stories of the JE corpus, ignoring filler segments. The correctness and accuracy scores obtained using only the top-level topic are shown in Table 4, assuming manually segmented stories. Topic accuracy is defined as the ratio between the number of correct detections and the total number of topics, and topic correctness as the ratio between the number of correct detections minus false detections (false alarms) and the total number of topics. The results for lower levels are very dependent on the amount of training material in each of these lower level topics (the second level includes over 1600 topic descriptors, and hence very few material for some topics).

When using topic models created with the non-rejected keywords, we observed a slight decrease in the number of misses and an increase in the number of false alarms. We also observed a slight decrease with manual transcriptions, which we attributed to the fact that the topic models were built using ASR transcriptions.

APP	ASR	Correctness	Accuracy
Manual	Manual	91.5	91.3
Manual	Auto w/o conf	93.8	91.5
Manual	Auto w/ conf	94.1	91.7
Auto	Auto w/ conf	93.9	91.4

Table 4. Topic indexation results.

These results represent a significant improvement over previous versions [17], mainly attributed to allowing multiple topics per story, just as in the manual classification. A close inspection of the table shows similar results for the topic indexation with auto or manual APP. The adoption of the word confidence measure made a small improvement in the indexation results, mainly due to the reduced amount of data to train the topic models. The results are shown in terms of topic classification and not story classification.

Whereas one could find comparable results for topic segmentation in the TDT2001 evaluation program [14], the topic indexation task has no parallel, because it is thesaurus-oriented.

7. CONCLUSIONS AND FUTURE WORK

This paper studied the impact of earlier errors in the last modules of our pipelined BN media watch system. This impact is in our opinion an essential diagnostic tool for its overall improvement.

Our APP module has a good performance, while maintaining a very low latency for stream-based operation. The impact of its errors on the ASR performance is small

(0.7% absolute) when compared with the manual references. The greatest impact of APP errors is in terms of topic segmentation, given the heuristically-based approach that is crucially dependent on anchor detection precision.

Our ASR module also has a good performance, although the results for European Portuguese are not yet at the level of the ones for languages like English, where much larger amounts of training data are available. We believe that unsupervised training approaches will be very helpful in this context. Our current work in terms of ASR is focused on dynamic vocabulary adaptation, and processing spontaneous speech, namely in terms of dealing with disfluencies and sentence boundary detection.

The ASR errors seem to have very little impact on the performance of the two next modules, which may be partly justified by the type of errors (e.g. errors in function words and in inflected forms are not relevant for indexation purposes).

Topic segmentation still has several pitfalls which we plan to reduce for instance by exploring video cues. In terms of topic indexation, our efforts in building better topic models using a discriminative training technique based on the conditional maximum likelihood criterion for the implemented Naive Bayes classifier [18] have not yet been successful. This may be due to the small amount of manually topic-annotated training data.

In parallel with this work, we are also currently working on unsupervised adaptation of topic detection models, unsupervised training of better acoustic models and improving speaker clustering by using the latest state of the art speaker identification techniques.

8. REFERENCES

[1] H. Meinedo and J. Neto, "A stream-based audio segmentation, classification and clustering preprocessing system for broadcast news using ann models," in *Proc. Interspeech '2005*, Lisbon, Portugal, 2005.

[2] M. Siegler, U. Jain, B. Raj, and R. Stern, "Automatic segmentation, classification and clustering of broadcast news," in *Proceedings DARPA Speech Recognition Workshop*, 1997.

[3] X. Zhu et al., "Combining speaker identification and BIC for speaker diarization," in *Proc. Interspeech '2005*, Lisbon, Portugal, Sept. 2005.

[4] J. Zibert et al., "The COST278 broadcast news segmentation and speaker clustering evaluation - overview, methodology, systems, results," in *Proc. Interspeech '2005*, Lisbon, Portugal, Sept. 2005.

[5] H. Meinedo, D. Caseiro, J. Neto, and I. Trancoso, "Audimus.media: a broadcast news speech recognition system for the european portuguese language," in *Proc. PROPOR '2003*, Faro, Portugal, June 2003.

[6] M. Mohri, F. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," in *ASR 2000 Workshop*, Sept. 2000.

[7] D. Caseiro and I. Trancoso, "A specialized on-the-fly algorithm for lexicon and language model composition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1281–1291, July 2005.

[8] D. Williams, *Knowing what you don't know: roles for confidence measures in automatic speech recognition*, Ph.D. thesis, Univ. Sheffield, UK, 1999.

[9] Adam L. Berger, Stephen Della Pietra, and Vincent J. Della Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[10] J.-L. Gauvain, L. Lamel, and G. Adda, "Developments in continuous speech dictation using the ARPA WSJ task," in *Proc. ICASSP '1995*, Detroit, USA, May 1995.

[11] E. Shriberg, "Spontaneous speech: How people really talk, and why engineers should care," in *Proc. Interspeech '2005*, Lisbon, Portugal, Sept. 2005.

[12] S. Matsoukas et al., "The 2004 BBN 1XRT recognition systems for english broadcast news and conversational telephone speech," in *Proc. Interspeech '2005*, Lisbon, Portugal, Sept. 2005.

[13] R. Barzilay, M. Collins, J. Hirschberg, and S. Whittaker, "The rules behind roles: Identifying speaker role in radio broadcast," in *Proc. AAAI 2000*, Austin, USA, July 2000.

[14] NIST Speech Group, "The 2001 topic detection and tracking (TDT2001) task definition and evaluation plan," 2001.

[15] I. Trancoso, J. Neto, H. Meinedo, and R. Amaral, "Evaluation of an alert system for selective dissemination of broadcast news," in *Proc. Eurospeech '2003*, Geneva, Switzerland, Sept. 2003.

[16] A. Gelbukh, G. Sidorov, and A. Guzmán-Arenas, "Document indexing with a concept hierarchy," in *Proc. NDDL 2001 - 1st Int. Workshop on New Developments in Digital Libraries*, Setúbal, Portugal, July 2001.

[17] R. Amaral and I. Trancoso, "Improving the topic indexation and segmentation modules of a media watch system," in *Proc. ICSLP '2004*, Jeju, Korea, Oct. 2004.

[18] C. Chelba, M. Mahajan, and A. Acero, "Speech utterance classification," in *Proc. ICASSP '2003*, Hong Kong, Apr. 2003.