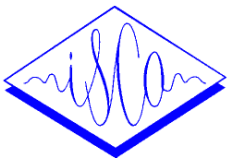

International Symposium on Chinese Spoken Language Processing (ISCSLP 2002)
Taipei, Taiwan
August 23-24, 2002

# RAPID PROTOTYPING AN OPERATOR ASSISTED CALL ROUTING SYSTEM

*Chun-Jen LEE* [1, 2]
[1] Telecommunication Labs.,
Chunghwa Telecom Co., Ltd., Chung-Li
cjlee@cht.com.tw

*Jason S. CHANG* [2]

[2] Department of Computer Science,
National Tsing Hua University, Hsin-chu
jschang@cs.nthu.edu.tw

## ABSTRACT

A prototype system to assist call routing task for telephone operators is reported in this paper. The system was developed based on a company organization profile with description of its divisions instead of a corpus of recorded and transcribed call-routing dialogs. An acoustic module and an information retrieval module were built specifically for this task. By integrating acoustic and information retrieval module, we have built a system with a satisfactory performance and provide a promising approach to call routing. Simulation results indicate that the proposed algorithm can improve call routing performance over baseline classification methods. A working system based on the proposed approach has been implemented and experimental results are presented.

## 1. INTRODUCTION

We proposed an end-to-end system for operator assisted call routing. The system was designed based on the idea that a telephone operator is capable to capture the intent of caller's inquiry but the operators may not be familiar or kept up to date with all the business of the organization. Therefore, he or she can then operate the call router via a speech interface. The system transforms the natural speech, matches the information related to the organization's directory, and finally determines the routing destination. The output is then displayed on the screen for the operator to carry out the actual routing action. The goal is to assist the operator in selecting the desired destination that matches caller's intension. This will expedite operator's response since there is no need to search a printed directory or enter the text query in Chinese.

Numerous previous works on call routing has been reported [11] [8] [4]. Most approaches in the literature require a corpus of routed calls to train a routing matrix or language model. Such corpora of dialog are sometimes difficult to obtain due to the concerns over invasion of privacy right. In addition to the acoustic module, a call router also needs to classify caller's request according to the routing destination. This task of classification is similar to text categorization in information retrieval (IR) or topic identification in speech research. An alternative approach is to adopt a form-based scheme which is commonly used in spoken understanding systems [5] [7] [3] [10]. A form consists of a set of keywords all (or most) of which need to be present in the caller's request in order for the destination to be activated. There are problems with adopting a form-based approach. Firstly, the acoustic module produces incorrect output, which may cause the IR module to fail. If we set the precision of the acoustic output too high, there could be insufficient information for the IR module to produce a routing destination. On the other hand, if the precision is set low for better recall, there could be too much noise causing the IR module to produce more than one destination. The second problem is manual effect needed to construct forms for each destination each time a new call router is built. Furthermore, one needs to provide synonyms to these independent keywords in the form. Otherwise, the system will be very fragile when faced with alternative ways of expressing the same routing request.

To cope with the problems mentioned above, we design the proposed call router with the following considerations:

- Automatic discovery of keywords and similar words.
- Automatic construction of forms.
- Tighter integration of acoustic and IR module.

To easy the burden of developing forms for destination and ensure consistency and coverage, we have experimented with a new approach [9] to extract keywords and their corresponding similar words for this specific routing task. Then, a collection of forms can be generated semi-automatically based on the description of destinations. By integrating acoustic and IR module, the system achieves satisfactory performance and provides a promising approach to call routing. An application developer is allowed to fine tune them in this framework.

## 2. TASK OVERVIEW

A speech query goes through speech recognizer and transforms into a keyword lattice. Then, call-routing classification is applied to generate a set of destination hypotheses with corresponding scores to be described latter. The scores are then compared with a predefined threshold. If more than one candidate destination pass the threshold, a disambiguation module should be invoked. On the contrary, if none of candidates passes the threshold, the request is rejected.

To avoid imposing on privacy of callers, no actual dialogue between a caller and a human operator has been recorded. The system was developed based on the organization profile of the Directorate General of Telecommunications (DGT) in Taiwan. The system consists of an acoustic module and an IR module built specifically for this task. A generic ASR engine was used as the kernel of the acoustic module without adapting to the application domain. The construction of IR module was based on term extraction and thesaurus discovery processes.

After field study of operators' routing task, we roughly classified the calls into three classes, destination name, activity, and indirect request in a way similar to the classification in [4]. We also observe that most calls can be classified according to destination name or activity. Hence, we focus on the classes of destination name and activity.

In the DGT call routing task, there are six routing destinations, including *General Planning Department, Public Telecommunications Department, Dedicated Telecommunications Department, Radio and TV Broadcast Technology Department, Radio Wave Regulatory Department,* and *Legal Office*. The organization profile of DGT describes the main responsibilities of its six departments. Parts of the missions of *GPD* are showed in Table 1.

Table 1: Mission of the General Planning Department (GPD).

| | |
|---|---|
| 1 | Drafting and supervising over the implementation of telecommunications policies. |
| 2 | Planning, promotion and supervision over the administration and cooperation of international telecommunications. |

For each of the main responsibilities of DGT departments, a form is formulated. For instance, the destination *GPD* can be represented by a number of forms representing its missions as follows:

FORM: GPD-1

| | | | |
|---|---|---|---|
| 1 | Telecom | | |
| 2 | Strategy | | |
| 3 | Drafting | Supervising | Implementation |

FORM: GPD-2

| | | | |
|---|---|---|---|
| 1 | International | | |
| 2 | Telecom | | |
| 3 | Planning | Promotion | Supervision |

The logical relation between slots in a form is represented by the AND operator and the relation between various admissible values in a slot is represented by the OR operator. In total, there are 95 forms constructed and 157 terms extracted from DGT profile.

## 3. ACOUSTIC MODELING AND UTTERANCE VERIFICATION CONVENTIONS

### 3.1 Task-independent Acoustic Modeling

A phonetically rich text database (TDB) was designed for target independent application. The speech database, consists of 399 phrases and short paragraphs that are chosen from TDB and read by 221 speakers via microphones. A total of 88,179 utterances were used for training acoustic models. All speech data was sampled and digitized at 16 kHz and pre-emphasized using a first order filter with a coefficient of 0.97. The samples were blocked into overlapping frames of 24 milliseconds in duration, where the overlap was set to 12 milliseconds. Each feature vector consisted of 24 features that included the 12 cepstral coefficients and 12 differenced cepstral coefficients extracted from a single frame of speech.

Typical hidden Markov model (HMM) based wordspotting systems were based on whole word models, and required training data for each of the keywords from a large database of speakers. They are thus appropriate in a small number of keyword task. Instead, in our keyword spotter system, keywords are represented by subword models. This has several potential advantages: a) support for very large keyword vocabularies, b) robustness with keyword variants that do not appear in training set, and c) portability of changing vocabulary without retraining acoustic models.

To reduce the likelihood computation, subsyllabic units, syllable initials and syllable finals, were used as basic HMM subword models [2]. Each initial and final model has 3 and 5 states, respectively. Overall there are 440 states for all the subsyllabic HMMs. A left-to-right HMM geometry with no state skip was chosen for all the models. A maximum of four mixture components per state is used. For rapid porting to applications, filler models also used the same models as the keyword models. An unconstrained network of syllabic units with penalty weights forms the filler models.

### 3.2 Utterance Verification and Confidence Measures

An utterance verification (UV) module is incorporated as a postprocessor of ASR to detect possibly erroneous keywords that have low confidence scores. A set of confidence measures [6] are applied and evaluated latter.

For every subword unit $u$ in a word candidate, a verification score is computed and defined as

$$LR_u = \frac{P(O \mid H_0)}{P(O \mid H_1)} = \frac{P(O \mid \lambda_u^c)}{P(O \mid \lambda_u^a)} \quad (1)$$

where $O$ is the observed speech segment, $H_0$ is the *null hypothesis* that subword unit $u$ exists in the segment of speech $O$, $H_1$ is the *alternative hypothesis* that subword unit $u$ is not present in the segment of speech $O$, and $\lambda_u^c$ and $\lambda_u^a$ are the corresponding subword and *anti-subword* HMM models for subword $u$, respectively. Then, a log likelihood ratio $LLR_u$ which takes the logarithm of $LR_u$ and normalizes it by the duration $l_u$ of $O$ is defined as

$$LLR_u = \{\log P(O \mid \lambda_u^c) - \log P(O \mid \lambda_u^a)\} / l_u . \quad (2)$$

Suppose a detected keyword is composed of $N$ subwords, the confidence measure ($CM$) for the keyword can be defined as a function of their log likelihood ratios.

$$CM = f(LLR_1, LLR_2 ..., LLR_N). \quad (3)$$

Several functional forms of the confidence measure is defined as following:

$$CM_1 = \frac{1}{L}\sum_u (l_u * LLR_u) \quad (4)$$

$$CM_2 = \frac{1}{N}\sum_u LLR_u \quad (5)$$

$$CM_3 = \frac{1}{N}\sum_u \frac{1}{1+\exp(-\alpha * LLR_u)} \quad (6)$$

where $L$ is total duration of the keyword, $CM_1$ is based on frame duration normalization, $CM_2$ is based on subword segment-based normalization, and $CM_3$ uses a sigmoid function to limit the dynamic range of the confidence measure.

## 4. CALL-ROUTING CLASSIFICATION

The difficulties of call-routing classification are due to insufficient or ambiguous information provided by the operators. Moreover, even when users utter an exact and unambiguous speech query, the correct routing destination may still be difficult to obtain due to imperfect speech

recognition. Therefore, it is necessary to include a ranking function to measure how close user's request is to one of the destinations. The similarity *P(D|Q)* between the query *Q* and the form, i.e. destination, *D* is formulated by combining the acoustic likelihood scores of all matching terms within-form provided by the HMM keyword spotter and information retrieval relevance score of form *D* to query *Q* provided by document retrieval formula. In our experiments, two scoring functions *SF1* and *SF2* as well as their corresponding weighting factors $\lambda_1$ and $\lambda_2$ were incorporated into Eq. (7) to tune the likelihood function of *P(D|Q)*. The ranking function can be formulated as following:

$$P(D \mid Q) = \lambda 1 \times SF1(AcousticScore) + \lambda 2 \times SF2(IRScore) \quad (7)$$

where *SF1* and *SF2* are the scoring functions of *AcousticScore* and *IRScore* respectively.

The acoustic score is formulated as follows:

$$AcoustiScore = \frac{1}{dl}\sum_{i=1}^{N} \log(p_{HMM}(kw_i)) \quad (8)$$

where

*N* is the number of matching terms between a form and a query,

*dl* is the form length (number of terms in a form),

$p_{HMM}(kw_i)$ is acoustic likelihood score of matching term $kw_i$.

We adopt the probabilistic approach [1] to formulate information retrieval score:

$$IRScore = -3.51 + 37.4 * X1 + 0.330 * X2 - 0.1937 * X3 + 0.929 * X4 \quad (9)$$

with

$$X1 = \frac{1}{\sqrt{N} + 1}\sum_{i=1}^{N} \frac{dtf_i}{dl + c_1}$$

$$X2 = \frac{1}{\sqrt{N} + 1}\sum_{i=1}^{N} \log \frac{qtf_i}{ql + c_2}$$

$$X3 = \frac{1}{\sqrt{N} + 1}\sum_{i=1}^{N} \log \frac{ctf_i}{cl}$$

$$X4 = N$$

where

*N* is the number of matching terms between a form and a query,

$qtf_i$ is the within-query frequency of the *i*th matching term,

$dtf_i$ is the within-form frequency of the *i*th matching term,

$ctf_i$ is the occurrence frequency in a collection of the *i*th matching term,

*ql* is the query length (number of terms in a query),

*dl* is the form length,

*cl* is collection length, i.e. the number of occurrences of all terms in our collection,

$c_1$ and $c_2$ are two constants.

## 5. EXPERIMENTAL RESULTS

An experimental study was performed to evaluate the effectiveness of the proposed approach. The section is composed of two parts. First, several functional forms of the confidence measure based on likelihood ratio scores are compared. Then, the integration of combining the scores from acoustic and IR module is investigated.

### 5.1 Comparison of UV measures

A comparison of the three different confidence measures given in Eq. (4)-(6) was undertaken. Performance is demonstrated both in terms of the receiver operating characteristic (ROC) curves and the sum of type I and type II errors plotted against the decision threshold settings. The sigmoid weight in Eq. (6) is parameterized using α=0.5. The likelihood of the *anti-subword* model, $\lambda_u^a$, for subword *u* was approximated by summing the average values of competing frame-level likelihood scores of the corresponding speech segment.

Fig. 1 shows comparison of the confidence measures with the ROC curve. Fig. 2 illustrates the sum of type I and type II error curves depending on the threshold values. The results of UV experiments indicate that the confidence measure $CM_2$ achieves the best performance of both low total errors and robustness to the setting of the confidence threshold in this work. The error rate does not change significantly over a range of threshold values. Therefore, the confidence measure $CM_2$ was used for UV in the subsequent experiments.

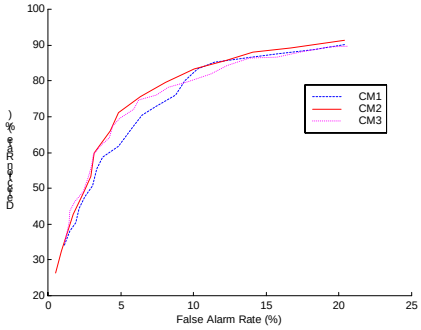


Figure 1: ROC curve.

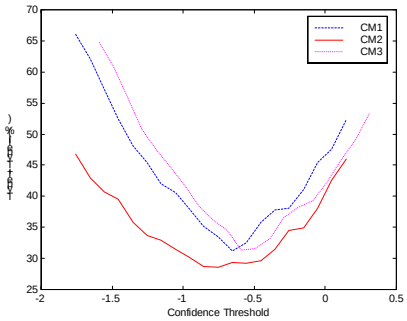


Figure 2: Error rate vs. confidence threshold.

### 5.2 Evaluation of the Task

The evaluation of the system so far has been limited to a laboratory setting. For the purpose of detailed comparisons, we classified the test utterances, i.e. queries, into short queries and long queries. In the short query experiment, the operator uttered only a sequence of telegraphic key phrases. In the long query experiment, the operator was allowed to utter spontaneous inquiries, which might at times be ungrammatical.

As we mentioned above, we ignored the call type of indirect request in our experiments. We recorded 131 utterances with 91 short queries and 40 long queries. Some query examples are given in the following:

第一類特許證。

(First class licence.)

幫我轉接至綜合規劃處。

(Please transfer me to *General Planning Department*.)

負責電信編碼規劃之部門。
(Department in charge of telecom coding and planning.)
要如何申請電信器材的審驗？
(How to apply for inspection of telecom equipment?)

In order to see whether combining the scores from acoustic and IR module produce the intended effectiveness, we conducted several experiments and compared the results. In the first three experiments, we skip the utterance verification module. Only a loosely predefined threshold of the HMM likelihood score is applied to filter word candidates. In the first experiment, we set λ2 to zero, which means ignoring IR score in Eq. (7). In the second experiment, we set λ1 to zero ignoring acoustic score in Eq. (7). The third experiment, non-zero acoustic and IR scores were used.

In the first three experiments, we obtain poor performance. It can be explained by that we employ a generic ASR engine without any fine-tuning, so there are a lot of non-relevant keywords included in the ASR candidate set. Then, to investigate the relative performance improvement, the utterance verification module with CM2 in Eq. (5) was integrated into the proposed framework. For both short and long queries, all the results indicate that the system with UV procedure outperforms the one without UV. It is effective for error reduction, as shown in Table 2. Evidently, the framework of integrating acoustic and IR score can achieve more improvement on the routing performance consistently.

Table 2: Routing error rates in various experimental setting.

| Experiments | Query types | Accuracy rate | |
|---|---|---|---|
| | | Top 1 | Top 2 |
| Experiment 1 ASR | Short query | 78% | 97% |
| | Long query | 65% | 80% |
| Experiment 2 IR | Short query | 58% | 85% |
| | Long query | 60% | 75% |
| Experiment 3 ASR+IR | Short query | 81% | 96% |
| | Long query | 68% | 80% |
| Experiment 4 ASR+UV | Short query | 93% | 97% |
| | Long query | 80% | 93% |
| Experiment 5 IR+UV | Short query | 85% | 88% |
| | Long query | 73% | 93% |
| Experiment 6 ASR+IR+UV | Short query | 94% | 97% |
| | Long query | 85% | 93% |

## 6. CONCLUSION

In conclusion, we have described the operator assisted call router for DGT in Taiwan. To build broad coverage and robust forms for linking callers' request and destination, we have exploited the statistical methods for discovery of keywords and synonyms. An effective way of integrating acoustic score and information retrieval score to classify call routing destination is also described. The preliminary evaluation shows that the approach is quite effective in routing calls in face of speech noise. The approach necessitates only minimal information and does not require a large call routing corpus. The system proved to be quite useful in providing one-shot question answering service to operators. For further studies, developing an additional dialogue module to facilitate the communication between the operator and the system for determining a unique routing destination is an important issue, especially as the need of porting the system to a telephone-based interface.

## 7. REFERENCES


[1] Chen, Aitao, Hailing Jiang and Fredric C. Gey. "Berkeley at NTCIR-2: Chinese, Japanese, and English IR Experiments," In *Proc. of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*, 2001, pp. 5-32 - 5-40.

[2] Chen, J.-K., F. K. Soong, and L.-S. Lee, "Large vocabulary word recognition based on tree-trellis search," In *Proc. ICASSP, Adelaide,* South Australia, April 1994, pp. II 137-140.

[3] Chu-Carroll, Jennifier. "Form-based reasoning for mixed-initiative dialogue management in formation-query systems," In *Proc. EUROSPEECH*, 1999, pp. 1519-1522.

[4] Chu-Carroll, Jennifier and Bob Carpenter. "Vector-based natural language call routing," *Computational Linguistics*, 25(3):361-388, 1999.

[5] Goddeau, D., Meng, H., Polifroni, J., Seneff, S., and Busayapongchai, S. "A Form-Based Dialog Manager for Spoken Language Applications" In *Proc. ICSLP,* Philadelphia PA, Oct, 1996, pp 701-705.

[6] Kawahara, T., C.-H. Lee, and B.-H. Juang. "Key-phrase detection and verification for flexible speech understanding," *IEEE Trans. on Audio and Speech Processing*, 6(6):558-568, 1998.

[7] Lamel, Lori.. "Spoken language dialog system development and evaluation at LIMSI," In *Proc. Inter. Symposium on Spoken Dialogue*, 1998, pp. 9-17.

[8] Lee, C.-H., R. Carpenter, W. Chou., J. Chu-Carroll, W. Reichl, A. Saad, and Q. Zhou. "A study on natural language call routing," In *Proc. Interactive Voice Technology for Telecommunications Applications*, 1998, pp. 37-42.

[9] Lee, C.-J. and J.S. Chang. "An operator assisted call routing system," In *Proc. of 16th Pacific Asia Conference on Language, Information and Computation*, 2002, pp. 271-280.

[10] Papineni, K. A., S. Roukos, and R. T. Ward. "Free-flow dialog management using forms," In *Proc. EUROSPEECH*, 1999, pp. 1411-1414.

[11] Riccardi, G., A. L. Gorin, A. Ljolje, and M. Riley. "A spoken language system for automated call routing," In *Proc. ICASSP*, 1997, pp. 1143-1146.