

TR-S-0026

ATR Chinese Speech Recognition System

張 樹武
ZHANG, Shuwu

張 勁松
ZHANG, Jinsong

2001,5

This report describes the details about building ATR Chinese speech recognition prototype and V1.0 system.

©ATR 音声言語通信研究所

©ATR Spoken Language Translation Laboratories

Contents

1	Introduction	2
2	The Characteristics of Chinese Language	3
3	The Prototype System of Chinese Speech Recognition	4
3.1	Chinese Speech and Language Corpus used in the Prototype System . . .	4
(3.1.1)	Chinese Speech Database	4
(3.1.2)	Chinese Language Materials	4
3.2	Acoustic Modeling	5
3.3	Language Modeling	5
3.4	Performance Tests	5
4	CAS00 Database and ATR Chinese Speech Recognition System V1.0	7
参考文献		9
付録 A	Where are models	10
A.1	The Prototype System	10
A.2	ATR Chinese Speech Recognition System V1.0	10
付録 B	config file used for demonstrations	10
B.1	config file used for Openhouse2000	10
B.2	config file used for HFSC workshp	12
付録 C	A Preliminary Investigation of Sub-Syllabic Modeling for Chinese Speech Recognition [1]	14
C.1	Prospective Chinese Sub-syllabic Sets	14
C.2	A Comparison of Sub-Syllabic Sets with Phonemes, Initials/Finals, and Premes/Postmes	15
(C.2.1)	Experimental setups and results	15
(C.2.2)	Remarks	15
付録 D	Some J-E-C Aligned Samples of Travel Conversations on Hotel Reser- vation Task	16

1 Introduction

For an aim at implementing real universal speech communication, we are attempting to enable ATRMATRIX to handle Chinese language. This work started from March 2000. We primarily scheduled to develop Chinese speech recognition module with three steps (as shown in Figure 1).

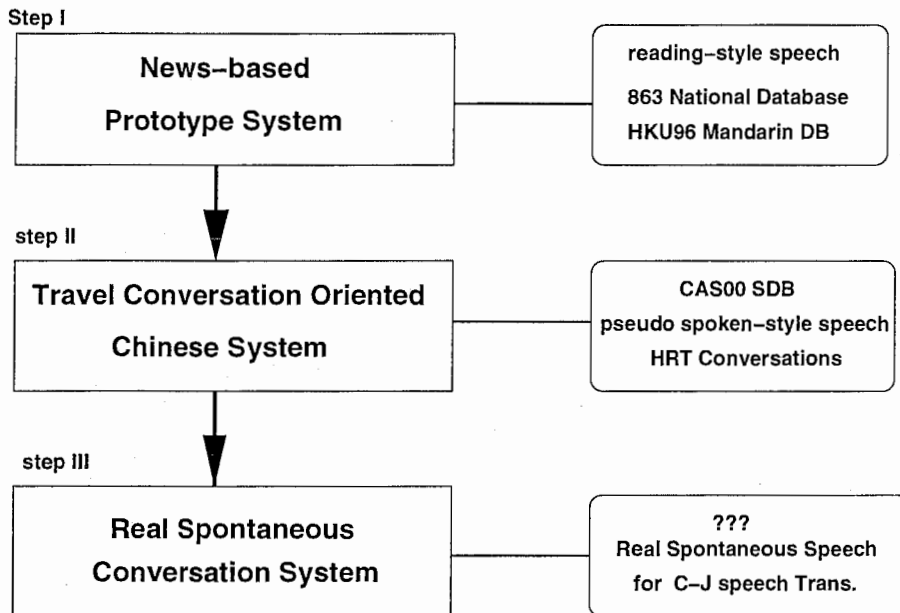


Figure 1: stages for the development of Chinese Speech Recognition Module for MATRIX

By now, we have established a general prototype Chinese speech recognition system and a travel conversation oriented system V1.0. They have been demonstrated respectively in ATR Research Exposition 2000 and International Workshop on Hand-Free Speech Communication (HSC) 2001. Meanwhile, we also trained three sets of models, news-based model, hotel reservation oriented model, and a hybrid model, for different purposes in research and demonstration. Now, we are working towards advanced topic: **real spontaneous speech communication**.

In this report, we describe the details about implementing ATR Chinese speech recognition prototype and V1.0 system.

2 The Characteristics of Chinese Language

Chinese is known as a mono-syllabic tonal language, in which each morpheme is uttered in the form of a mono-syllable associated with a specific pitch tone. The total number of phonetically distinguishable syllable (including tone distinction) is around 1300, and the number of base syllables which only differs in segmental structure (not tone distinction) is around 410. [1] [2] [3]

From the view of segmental events, one base syllable may consist of at most four units: consonant, glide, nucleus and coda. Among of them, only the nucleus is obligatory and other three are optional. According to different principle of partition, these segmental events can be expressed by different acoustic unit sets in speech recognition. They are

- 36 phoneme set
This is phonetics-based classification. The phoneme set consists of 22 consonants and 14 vowels.
- 60 Initials/Finals
This is a sort of traditional decomposition. It divides a base syllable in the front of glide, that is the vowel in a base syllable is comprised with glide, nucleus and coda. This set includes 22 Initials and 38 Finals.
- 78 Premes/Postmes
This is a sort of balanced decomposition, especially for speech recognition. It divide a syllable in the back of glide so that both Preme and Postme in a syllable could be balanced in length. This partition includes 54 Premes and 24 Postmes.

Chinese words are variable in the length of syllable combination. In general, word with less than five syllables can cover around 96% of Chinese vocabulary.

In addition, since it is hard to model tone variations in connected words or continuous speech, usually, tone information has been ignored in most Chinese speech recognition systems.

3 The Prototype System of Chinese Speech Recognition

3.1 Chinese Speech and Language Corpus used in the Prototype System

(3.1.1) Chinese Speech Database

In prototype system, we used two general Chinese speech database. The first one is HKU96, which made by The Hong Kong University in 1996. Another one is the Chinese national speech database for speech recognition (abbreviated as “863”DB).

SDB	Type	Recording Env.	Amount of Speakers	Num. of Utterances
HKU96	Reading	16k Sampling,16bits Quan.	20 SPKs.	53,751
863DB	Reading	16k Sampling,16 bits Quan.	166 SPKs.	96,745

Table 1: Description of Chinese Speech Data Bases

Meanwhile, we also recorded some conversational speech based on ATR hotel reservation task, HRT for short, with two speakers in ATR (Shuwu Zhang & Jinsong Zhang) for the purpose of adaptation and testing.

(3.1.2) Chinese Language Materials

Since we have not enough Chinese language corpus corresponding to ATR domain to date, we are tentatively, using a kind of trade-off combined corpus to train language model in the prototype system. The combined corpus consists of Parts of “People’s Daily” News corpus, the transcriptions of speech DBs, parts of JEK sentences related to travel conversations, and some machine and manual translated HRT dialogues (see table 2). It includes about 300k tokens. The vocabulary occurred in this combined corpus is 9183 words.

LDB	Type	Num. of Utterances.
parts of JEK set	Spoken	1000
25 MT Dialogues	Conversation	930
5 HT Dialogues	Conversation	237
Common used Utts		45
People’s Daily	News	1430
In total		3597

LDB: Language Database, Utts: Utterances

MT: machine translated, HT: manual-translated

Table 2: Description of Chinese Language Corpus Used for the Prototype System

3.2 Acoustic Modeling

Based on HKU96 and 863 Databases, we trained a couple of context-dependent and speaker-independent Chinese acoustic models using ATRSPREC toolkit [4]. The basic configuration for acoustic modeling can be seen as follows.

```
Acoustic Units: Initials(22) + Finals(38) + Silence(1)
Parameter Dimensions: pow(1)+mfcc(12)+dpow(1)+dmfcc(12)
Model Type: Context dependent HMnet
              with 1000 states and 5 distributions for each state
Models: Gender dependent male/female models
Training Mode: Decision Tree aided MLSSS algorithm
Training Set: HKU(18 Speakers)+863(160 Speakers)
Testing Set: 863(6 Speakers)+ 2 speakers in ATR
```

3.3 Language Modeling

Since we don't have detailed language corpus for intensive language modeling by now, we just trained some word-based bigram models in the prototype system. By combing parts of writing-style "The People's Daily" News corpus with spoken-style travel arrangement conversations, we got a combined corpus with total of 300k tokens and a vocabulary including about 9,183 words and 9,804 pronunciation variants. Based on this combined corpus, we train a word-based bigram language model which is applicable in both dialogue and dictation occasions.

3.4 Performance Tests

With above acoustic and language modeling, some experimental results have been shown in table 3 and table 4.

Firstly, we partitioned 863 national database with 6 speakers (3 males and 3 females) as open speaker testing set. Since testing sets in either speech data or language materials are included in training sets. It is a sort of test with regard to model close & speaker open. As shown in table 3, the average accuracy for males can be achieved 90.60%. But for all of three females, the accuracy are totally poor. It would be partly because of the mismatch in speech style between training and testing. We, furthermore, conducted the test with a close female speaker to check how worse the female model was. Result showed that close speaker has a surprising high word accuracy up to near 96%. This result somehow may smooth above judgment with open speakers regarding the female model in some degree.

Meanwhile, we found that almost 40% errors for each test are insertion error. and most of this type of errors occurred in the beginning or end of the utterance. It may be mainly due to improper configuration in EPD function.

For meeting the expectation of ATR application, we further adapted the models with new speakers and switched them into new domain based on hotel reservation conversation by VFS-MAP [4]. The results in Table 4 showed a significant improvement with adaptation, and an efficient switch to a new domain with a small amount of speech data.

speakers	sets	acc/cor	total words	ins	del	sub	real time	num. of sents
Open Speakers								
M97	A	91.96/95.44	4688	163	27	187	0.83	613
M98	B	91.57/94.93	4756	160	27	214	0.82	626
M99	C	88.36/93.82	4871	266	19	282	0.76	631
M		90.60/94.72	14315	589	73	683	0.80	1870
F97	A	49.14/71.26	4367	966	124	1131	0.84	577
F98	B	57.58/76.59	4366	830	109	913	0.80	580
F99	C	64.18/80.31	4484	723	132	751	0.77	589
F		57.03/76.09	13217	2519	365	2795	0.80	1746
Closed Speaker								
F00	C	95.97/97.52	4068	63	14	87	1.17	520
New speaker and recording conditions								
szhang	A	87.17/94.04	1979	136	6	112	1.71	222

The tests were conducted with close models and open/close speakers

Table 3: Performance on 863 sets

speakers	sets	acc/cor	total words	ins	del	sub	real time	num. of sents
Without Adaptation								
szhang	863 News	87.17/94.04	1979	136	6	112	1.71	222
With Adaptation								
szhang	863 News	93.28/96.72	1979	68	3	62	1.35	222
Adapted to New Domain (HRT)								
szhang	hrt dialogues	92.44/97.36	1666	82	1	43	1.86	237
jzhang	hrt dialogues	81.19/90.59	2126	200	39	161	3.53	299

Table 4: Adapted models with new speakers/new domain

4 CAS00 Database and ATR Chinese Speech Recognition System V1.0

A project for collecting script-scheduled HRT-based speech data have been finished in cooperation with Chinese Academy of Sciences (CAS) in February, 2001. This database consists of 50 hotel reservation dialogues which is the subset of HRT-related SLDB corpus, 1000 supplement JEK sentences, and a group of balanced digit strings. Speakers are total of 100 including 50 males and 50 females. We, hereafter, call this database as CAS00 DB. Some detailed information about recording CAS00 database can be seen in following specification.

Notes on translating EJtext of 50 dialogues based on Hotel Reservation Task
into the Chinese text

- 1): For each utterance, there are both English and Japanese explanations.
You can select one you preferred to translate.
- 2): If there are a few explanations in English side (like E, E-, or E+),
E+ would be the best selection.
- 3): You can change parts of hotel names, person name or city places as easy-reading
or famous Chinese names. but please remain a small part unchanged.
And please let us know what have been changed.
- 4). Japanese name should be uttered according to Chinese pronunciations.
- 5): The translation style should be oriented to daily spoken usage.
- 6): Name convention should be the same title with suffix ".CTEXT" like:
(TAC22011.JETEXT => TAC22011.CTEXT)

Notes on script-scheduled speech data collection

- 1): reading text: 50 dialogues on HRT + 1000 utterances in JEK set
- 2): speakers: 100 (50 male + 50 female)
- 3): the distribution of speakers:
Beijing area (North China): 80, Shanghai area: 10, Hunan: 2, Jiangxi: 2,
Sichuan: 2, Guangdong: 2, Fujian: 2
- 4): each speaker:
2 conversations(2 clerk side and 2 customer side) on HRT +20 utterances in JEK
+ 5 continuous digits
- 5): 16kHz sampling, S/N > 50dB
- 6): Name convention:
HRT side: subdir/title.order.role.16k
here: order is a 4 digits with 10 as increment, Role: A:customer, B:clerk
(like the 4th utterance in TAC22011 dialogues should be named
TAC22011/TAC22011.0040.A.16k)
JEK side: utterance title.16k (i.g. B00664.16k)
Digits: DIGITS.xxxx.B.16K (i.g. DIGITS.0010.B.16k)

With this travel conversation oriented speech database, we have adapted general Chinese acoustic models and language models to specific ATR-task domain and successfully imple-

mented ATR Chinese speech recognition system V1.0 version. It has been demonstrated in International Workshop on Hand-Free Speech Communication (HSC) 2001.

参考文献

- [1] S. Zhang, J. Zhang, S. Nakamura, and Y. Sagisaka. A preliminary investigation of sub-syllabic modeling for chinese speech recognition based on HMnet. In *ASJ 2000 Fall Meeting*, pages 127–128, 2000.
- [2] J. Zhang, S. Zhang, Y. Sagisaka, and S. Nakamura. Modeling segmental context effect for chinese speech recognition. In *IEICE, Technical Report, Vol. 2001, No. 11*, pages 37–42, 2001.
- [3] J. Zhang, S. Zhang, Y. Sagisaka, and S. Nakamura. A hybrid approach to enhance task portability of acoustic models in chinese speech recognition. In *to appear in Eurospeech*, 2001.
- [4] ATR, Spoken Language Translation Research Laboratories. ATRSPREC user, programmer and reference manual. Technical report, ATR, 2000.

付録 A Where are models

A.1 The Prototype System

Acoustic Models /DB/PDB/PUB/AM/HMnetQCR.M.5.1000.hku+863.001.bin,
/DB/PDB/PUB/AM/HMnetQCR.F.5.1000.hku+863.001.bin,
/DB/PDB/PUB/AM/AM.M.M001.863.bin,
/DB/PDB/PUB/AM/AM.M.M002.863.bin
Lexicon: /DB/PDB/PUB/LM/creco_lex.1.cv
Language Model:/DB/PDB/PUB/LM/bigram.general.bin

A.2 ATR Chinese Speech Recognition System V1.0

Acoustic Models /DB/PDB/PUB/AM/AM.M.bin, /DB/PDB/PUB/AM/AM.F.bin,
/DB/PDB/PUB/AM/AM.M.M001.hrt.bin,
/DB/PDB/PUB/AM/AM.M.M002.hrt.bin
Lexicon: /DB/PDB/PUB/LM/LEX.cas+
Language Model:/DB/PDB/PUB/LM/bigram.atr.bin

付録 B config file used for demonstrations

B.1 config file used for Openhouse2000

/DB/PDB/PUB/DEMO/openhouse/config.chinese

```
# ATRSPREC r06r06 Config file for Chinese Recognition
# audio setting
Demo:adin=$SPREC_ADIN
Demo:daout=$SPREC_DAOUT
Demo:inputByteorder=$SPREC_INPUTBYTEORDER # input audio
Demo:infileByteorder=BigEndian # input file
Demo:machineByteorder=$SPREC_MACHINEBYTEORDER # input eventloop
Demo:outputByteorder=$SPREC_OUTPUTBYTEORDER # output audio
Demo:outfileByteorder=BigEndian # output file
I/Ocontrol:inputFormat=NoHeader
I/Ocontrol:inputParamSize=160
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NULL
I/Ocontrol:outputParamSize=26
I/Ocontrol:outputParamType=float
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:rpcNumber=2

ATRepd:SamplingFrequency=16000
ATRepd:energyThreshold=100
ATRepd:upperDispersionThreshold=50
ATRepd:lowerDispersionThreshold=10
ATRepd:orderInMs=300
ATRepd:alpha=1.06
ATRepd:skewInMs=300
ATRepd:s1TimerLimitInMs=200
ATRepd:epdFramesInMs=2000
ATRepd:framePointsInMs=7.5
ATRepd:inputByteOrder=none
ATRepd:sendNonspeech=0
```

```

ATRcpd:track=0

ATRwave2cep:Preemphasis=0.98
ATRwave2cep:FrameLength=20
ATRwave2cep:FrameShift=10
ATRwave2cep:SamplingFrequency=16000
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:LpcOrder=16
ATRwave2cep:CepstrumOrder=12
ATRwave2cep:FrequencyWarping=mel
ATRwave2cep:FilterBankOrder=16
ATRwave2cep:CutoffLowFrequency=0
ATRwave2cep:CutoffHighFrequency=8000
ATRwave2cep:AnalysisType=fft
ATRwave2cep:DebuggingLevel=0
ATRwave2cep:Subtract=logpow+cep

ATRcep2para:CepstrumOrder=12
ATRcep2para:LDA=
ATRcep2para:DeltaCepstrumWindow=9
ATRcep2para:deltaCepstrumPadding=zero
ATRcep2para:rho=1.0
ATRcep2para:OutputParameter=pow+cep(12)+dpow+dcep(12)

ATRlattice:lexicon=/DB/PDB/863/LM/sprec/creco.lex.1
ATRlattice:amname=/DB/PDB/863/AM/sprec/HMnetQCR.F.hku+863.001,
                /DB/PDB/863/AM/sprec/HMnetQCR.M.hku+863.001
ATRlattice:active_model=all
ATRlattice:lmscale=6.0,10.0
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=/DB/PDB/863/LM/sprec/bigram.general.bin
ATRlattice:active_lmodel=1
ATRlattice:FSA=
ATRlattice:beam=65,65
ATRlattice:work_area=300,50
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=sil
ATRlattice:dimension=26
ATRlattice:state_skip=OFF
ATRlattice:phone_boundary=ON
ATRlattice:word_boundary_skip=2
ATRlattice:word_merge=all
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
ATRlattice:UTT_END_delay=-1
ATRlattice:amscale=1.000000
ATRlattice:pause_beam_rate=1.0

ATRresult:N_best=1
ATRresult:N_best_out=/dev/null
ATRresult:Lattice_out=/dev/null

```

B.2 config file used for HSC workshp

/DB/PDB/PUB/DEMO/handfree/config.atr

```
# audio setting
Demo:adin=$SPREC_ADIN
Demo:daout=$SPREC_DAOUT
Demo:inputByteorder=$SPREC_INPUTBYTEORDER # input audio
Demo:infileByteorder=BigEndian # input file
Demo:machineByteorder=$SPREC_MACHINEBYTEORDER # input eventloop
Demo:outputByteorder=$SPREC_OUTPUTBYTEORDER # output audio
Demo:outfileByteorder=BigEndian # output file
I/Ocontrol:inputFormat=NoHeader
I/Ocontrol:inputParamSize=160
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NULL
I/Ocontrol:outputParamSize=26
I/Ocontrol:outputParamType=float
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:rpcNumber=2

ATRepd:SamplingFrequency=16000
ATRepd:energyThreshold=100
ATRepd:upperDispersionThreshold=50
ATRepd:lowerDispersionThreshold=10
ATRepd:orderInMs=300
ATRepd:alpha=1.06
ATRepd:skewInMs=300
ATRepd:s1TimerLimitInMs=200
ATRepd:epdFramesInMs=2000
ATRepd:framePointsInMs=7.5
ATRepd:inputByteOrder=none
ATRepd:sendNonspeech=0
ATRepd:track=0

ATRwave2cep:Preemphasis=0.98
ATRwave2cep:FrameLength=20
ATRwave2cep:FrameShift=10
ATRwave2cep:SamplingFrequency=16000
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:LpcOrder=16
ATRwave2cep:CepstrumOrder=12
ATRwave2cep:FrequencyWarping=mel
ATRwave2cep:FilterBankOrder=16
ATRwave2cep:CutoffLowFrequency=0
ATRwave2cep:CutoffHighFrequency=8000
ATRwave2cep:AnalysisType=fft
ATRwave2cep:DebuggingLevel=0
ATRwave2cep:Subtract=logpow+cep

ATRcep2para:CepstrumOrder=12
ATRcep2para:LDA=
ATRcep2para:DeltaCepstrumWindow=9
ATRcep2para:deltaCepstrumPadding=zero
ATRcep2para:rho=1.0
```

ATRcep2para:OutputParameter=pow+cep(12)+dpow+dcep(12)

ATRlattice:lexicon=/DB/PDB/PUB/LM/LEX.cas+
ATRlattice:amname=/DB/PDB/PUB/AM/AM.M.bin,
/DB/PDB/PUB/AM/AM.F.bin

ATRlattice:active_model=all
ATRlattice:lmscale=6.0,10.0
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=/DB/PDB/PUB/LM/bigram.atr.bin
ATRlattice:active_lmodel=1
ATRlattice:FSA=
ATRlattice:beam=65,65
ATRlattice:work_area=300,50
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=sil
ATRlattice:dimension=26
ATRlattice:state_skip=OFF
ATRlattice:phone_boundary=ON
ATRlattice:word_boundary_skip=2
ATRlattice:word_merge=all
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
ATRlattice:UTT_END_delay=-1
ATRlattice:amscale=1.000000
ATRlattice:pause_beam_rate=1.0

ATRresult:N_best=1
ATRresult:N_best_out=/dev/null
ATRresult:Lattice_out=/dev/null

付録 C A Preliminary Investigation of Sub-Syllabic Modeling for Chinese Speech Recognition [1]

C.1 Prospective Chinese Sub-syllabic Sets

As described in section 2, Chinese sub-syllabic sets can be basically classified into three categories according to the different partitions of the segmental structure - 36 phonetic phonemes, 60 Initials and Finals, or 78 Premes and Postmes. Following Tables (Table 5,6 and 7) enumerate the details of these three partitions.

Consonants (22)	b, p, m, f, d, t, n, l, z, c, s, zh, ch, sh, r, g, k, h, j, q, x, ng
Vowels (14)	a1(ba,wa), a2(an,ai), a3(ang,ao), e1(ge), e2(ei,ye), e3(en,eng), i1(bi,qi), i2(zi,ci), i3(zhi,chi), o1(bo), o2(ou), u, yv,er
Standard phonetic classification. i.g. bian => b + i1 + an	

Table 5: Basic Phonetic Classification - 36 Phonetic Phonemes

Initials (22)	b, p, m, f, d, t, n, l, z, c, s, zh, ch, sh, r, g, k, h, j, q, x, ng
Finals (38)	a, ai, an, ang, ao, e, en, ei, eng, er, i1, i2, i3, ia,ian,iang,jiao,ie,in, ing, iong, iu, u, ua, uan, uai, uang, ui, un, uo, ueng, ong, v, van, ve, vn, o, ou
segment a syllable before the glide part. i.g. bian => b + ian	

Table 6: Traditional Decomposition - 60 Initials/Finals

Premes (54)	b, p, m, f, d, t, n, l, z, z,c,s, zh, ch, sh r, g, k, h, j, q, x, ng, y, w, bi, pi, mi, di, ti, ni, li, ji, qi, xi, du, tu, nu, lu, zu, cu, su, zhu, chu, shu, ru, gu, ku, hu, nv, lv, jv, qv, xv, yv
Postmes (24)	a, ai, an, ang,ao, e, en, ei, eng, er, i1, i2, i3 in, ing, iu, u, ui, un, ong, v, vn, o, ou
segment a syllable after the glide part. i.g. bian => bi + an	

Table 7: Balanced Decomposition - 78 Premes/Postmes

C.2 A Comparison of Sub-Syllabic Sets with Phonemes, Initials/Finals, and Premes/Postmes

(C.2.1) Experimental setups and results

We conducted the experiments based on HKU96 Chinese Database. Acoustic feature vector consists of 12 MFCCs, their first-order derivations, power and delta power. Since numbers of sub-syllabic units are different in the three categories, there should be different levels of model sharing for them. However, in order to make clear the recognition performance of each category on the basis of same size of training data, we trained models with the same number of states of mixtures for all three categories. Each model has 1000 states with 5 mixtures per state. All the sub-syllabic units were modeled in the same length of 3 states. The language models used in this preliminary investigation are a sub-syllable based bigram and a syllable based bigram for comparing the influence with different levels of the language units. Table 9 showed the results of the two recognition experiments, by the SER standing for the syllable error rates, and SSER for sub-syllable error rates.

	Speakers	Num. of CS	Num. of IS	Others	In Total
Training Set	0m-8m,0f-8f	17,789	29,926	700	48,408
Testing Set	9m, 9f	1,943	3,306	60	5,309

CS: continuous sentences; IS: isolated syllables.
Others included a group of words, digit strings, and rhymed syllables .

Table 8: Database Partition

Sub-syllabic units	SSER	SER
38 Phonemes	43.50%	13.37%
60 Initials/Finals	38.96%	17.63%
78 Premes/Postmes	37.81%	16.02%
SSER: Sub-Syllabic Error Rate		
SER: Syllabic Error Rate		

Table 9: Preliminary recognition results based on three subsyllabic modeling approaches

(C.2.2) Remarks

From above results, we may see that for syllable recognition, the modeling based on the set of preme and postme outperforms the other two sets, under the present condition of limited training data. The phone-based method showed a better performance in the recognition of sub-syllabic units compared to other two sets, since the original labeling information of Initials/Finals and Premes/Postmes sets were converted from the labels of phonemes. It would be not enough to say that phone-based classification is certainly better than others by this preliminary investigation. Thus, the results based on this preliminary investigation should be further confirmed and renewed by conducting experiments on larger speech databases.

付録 D Some J-E-C Aligned Samples of Travel Conversations on Hotel Reservation Task

J|もしもし。

E|Hello.

C|喂

E|Thank you for calling Hotel New York reservations. |How may I be of service?

J|お電話ありがとうございます。|ニューヨークホテルの予約係でございます。|御用をお伺いします。

C|欢迎 你 来 电话 这里 是 纽约大厦 预约处

J|予約をお願いしたいのですが。

E|I'd like to make a reservation.

C|我 想 预订 住房

E|All right. |Now when will you need this room?

J|かしこまりました。|いつ、御宿泊の予定でございますか。

C|好的 预定 什么时候 住宿

J|九月二日<futsuka>の金曜日から三日間<mikkakan>お願いしたいのですが。

E|I'd like a reservation for three days, from Friday, September second.

C|从 九 月 二 号 星 期 五 起 三 天

E|All right. |And how many persons are in your party?

J|かしこまりました。|[え] 何名様でいらっしゃいますか。

C|好的 几位

J|わたしと妻の二人です。

E|Two of us. |Me and my wife.

C|两 位 我 和 我 妻 子

E|All right. |Do you have a particular preference for a room style or type?

J|[え] ではお部屋のスタイルもしくはタイプに御希望がございますか。

C|那么 对 房 间 的 款 式 或 类 型 您 有 要 求 吗

J|妻との銀婚式の記念ですので、

E|[well] Actually this is for our silver wedding anniversary

C|实际 上 这 是 为 了 我 们 的 银 婚 纪 念 日

J|できれば広くて雰囲気のよい部屋があればなと思ってます。

E|so I'd like to have a big room with a nice ambiance.

C|所以 最好 是 一 个 宽 敞 的 气 氛 不 错 的 房 间

E|All right.

J|分かりました。

C|知 道 了

E|[well] For such occasions we would [ah] highly recommend our suites.

J|[あー] そのような御希望でしたら、わたくしどものスイートをぜひお勧めいたします。

C|如 果 是 这 样 我 建 议 你 们 使 用 我 们 的 豪 华 客 房

J|ほかにも妻を喜ばせたいので、部屋で予算がオーバーするかどうか気になっているのですが、

E|I would like to please my wife so as long as the room doesn't go over the budget, that's okay.

C|我 想 让 我 妻 子 感 到 快 乐 只 要 房 费 不 超 过 预 算 就 行

J|スイートルームはおいくらでしょうか。

E|But how much would a suite cost?

C|那么 豪华 客房 的 费用 是 多少 呢

E|[well] Under normal circumstances [um] a double occupancy suite would go for eight hundred dollars a night.

J|[え、そうですね] [え] 通常、お二人様でスイートのお部屋一泊八百ドルで承っております。

C|通常 情况 下 一个 双人 套房 是 一 天 八百 美元

E|Now it's very large and luxurious

J|[え] 非常に広いお部屋で、[えー] 豪華なお部屋でございます。

C|这 是 一 个 相 当 大 而 且 豪 华 的 房 间

E|and it has a very large living room and a bedroom.

J|[えー] リビングと、[お] 広いベッドルームがございます。

C|有 起 居 室 和 宽 敞 的 卧 室

J|[ふうん] それだけで予算を全部使ってしまいます。|何か割引|はしていただけないでしょうか。

E|That sort of uses up my entire budget.|Couldn't I get some kind of a discount?

C|这 已 经 超 出 我 的 整 个 预 算 能 不 能 给 我 一 些 折 扣

E|[ah] You know you sure can.|We have a whopping discount

J|[そうですね] 大幅な割引|を御用意いたしまして、

C|确实 可 以 我 们 有 一 项 特 殊 优 惠

E|that will let you enjoy our luxurious room service.

J|その分、贅沢なルームサービスを、お楽しみいただけるかと思えます

C|它 将 使 您 享 受 我 们 的 超 豪 华 客 房 服 务

J|一体、いくら割引|いていただけますか。

E|Can you tell me exactly how much this discount you have is?

C|你 能 告 诉 我 确 切 的 减 价 幅 度 吗

E|With our weekend service and our Silver Anniversary discount,

J|[そうですね、え] わたくしどもの週末サービスと、それから銀婚式の割引|を付け加えさせていただきますら、

C|按 照 我 们 的 周 末 优 惠 和 银 婚 纪 念 折 扣

E|We'll take three hundred dollars off the regular price,

J|三百ドルお引きすることが、通常の[おー] レートからお引きすることができます。

C|我 们 将 从 正 常 价 格 中 减 价 三 百 美 元

E|Making it only five hundred dollars.

J|五百ドルでお泊まりいただけます。

C|仅 仅 五 百 美 元 一 天

J|週末割引|が四十パーセントのところもあるのですが、もう少し割引|いていただけませんか。

E|[well] There are some places that give a forty percent discount for their weekend service.

|Can't you take a little more off the price?

C|周 末 优 惠 百 分 之 四 十 的 地 方 也 有 能 不 能 再 便 宜 点 儿

E|Gosh, you got me there.

J|分かりました。

C|知 道 了

E|Tell you what I'll do.|[um] I'll set up for you to have free room service breakfast.

J|ではこのようにいたしましょう。|[えー] ルームサービス [うー] で [え] 御朝食をサービスさせていただきます。

J|それをお願いします。

E|Okay, I'll take it.

C|那就 这样 吧

E|Terrific.|Could I have your full name and your credit card number please?

J|ありがとうございます。| [う] ではお名前をフルネームで、それとクレジットカードの番号をお願いいたします。

C|谢谢 那么 请 让 我 知道 您 的 全名 和 信用卡 号码

J|はい、鈴木和夫と申します。

E|Yes, my name is Kazuo Suzuki.

C|好 我 叫 铃木 和夫

J| [え] カードは、ビザで、

E|And my card is a VISA card

C|信用卡 是 维萨卡

J|四八八三、五八零<rei>零<rei>、四零<rei>八八、一七一八です。

E|number four eight eight three five eight zero zero four zero eight eight one seven one eight.

C|号码 是 四 八 八 三 五 八 零 零 四 零 八 八 一 七 一 八

E|All right, Mr. Suzuki.|Let me repeat that for you.

J|では復唱いたします、鈴木様。

C|好的 铃木 先生 让 我 重复 一下

E|It's Mr. Kazuo Suzuki.

J|鈴木和夫様で、

C|鈴木 和夫 先生

E|Your VISA card number four eight eight three five eight o o four o eight eight one seven one eight.

J|ビザの番号が、四八八三、五八零<zero>零<zero>、四零<zero>八八、一七一八番でございますね。

C|您 的 维萨卡 号码 是 四 八 八 三 五 八 零 零 四 零 八 八 一 七 一 八 对吗

E|Now, I'm reserving a suite for you

J| [え] では予約の確認をいたします。

C|现在 我 正 为 您 预订 一个 豪华 套房

E|for three nights from Friday September second

J|九月二日<futsuka>金曜日より、三日間<mikkakan>で、

C|从 九 月 二 号 星 期 五 开 始 共 三 天

E|and that'll be at five hundred dollars per night, excluding tax, but including the free breakfasts.

J|こちらは税別で、 [え] それから朝食をサービスでお付け、させていただきまして、

五百ドルのお部屋を予約させていただきました。

C|价格 是 每 天 五 百 美 元 不 包 括 税 金 另 外 提 供 免 费 早 餐

J|よろしくお願ひします。

E|Thank you very much.

C|非常 感谢

E|No problem.|Now my name is John Phillips at the reservation desk.

J|もちろんでございます。|ありがとうございます。

|わたくしは予約担当、 [え] (フィリ) ジョン・フィリップスでございます。

C|没 问题 我 是 前 台 的 约翰 菲利普斯

E|If you have any questions please feel free to give me a call at any time.

川何なりとご質問がございましたら、お電話で御連絡いただきますようお願いいたします。

|本日はありがとうございました。

c|如果 您 有 什 么 问 题 请 随 时 给 我 打 电 话 联 系

HOTEL RESERVATION: TAS32015

E|Queen Mary, reservations.|May I help you?

J|はい、クイーンメアリー号予約係<kakari>でございます

C|这里 是 玛丽 女王 号 预约处

J|ちょっと、お尋ねしますが、

E|I have a question for you.

C|我 想 问 一 个 问 题

J|そちら、ホテルのように宿泊できるんでしょうか。

E|I'm wondering if I can stay on your ship just like I'd stay at a hotel?

C|我 想 知 道 ， 我 能 像 住 在 饭 店 一 样 住 在 你 们 的 客 船 上 吗 ？

J|停泊中の豪華客船に泊まると聞いてですね、ちょっと興味があるものですから。

E|I was told that I can stay on your luxury liner while in port.

C|听 说 你 们 的 豪 华 客 船 在 停 泊 时 可 以 住 宿

E|Sure, would you like to make a reservation?

J|はい、御宿泊を御希望でらっしゃいますか。

C|是 的 ， 您 想 预 订 吗 ？

J|ええ、そうなんです、ぜひあしたにでも空いていたらと思ひましてねえ。

E|Yes, and if possible I like to stay tomorrow.

C|对 ， 假 如 可 能 我 想 明 天 住 宿

E|Sure, we still have a variety of cabins available.

J|はい、今でしたらまだ予約のほうに空きがございますので、

(いろいろな) いろいろなキャビンをお選びになって、お泊まり願うことができますよ。

C|当 然 可 以 ， 我 们 仍 然 有 各 种 类 型 的 客 舱

E|You can get the feel of a voyage.

J|ちょうど船旅のような気分を味わっていただくことができます。

C|您 能 领 略 到 航 海 的 感 觉

J|それはいいですね、じゃあ旅をしているお客さんとおなじようにいろんな施設、利用できるわけですね。

E|That's great.|So we can use all the facilities just like the rest of the passengers, right?

C|太 好 了 ， 那 么 我 能 像 别 人 一 样 使 用 所 有 的 游 乐 施 施 吗 ？

E|You can use the restaurants, bars, nightclubs and most of the entertainment facilities.

J|はい、レストラン、バー、ナイトクラブ、そのほか、アミューズメント施設はほとんど御利用願えますよ。

C|您 能 使 用 餐 厅 酒 吧 夜 总 会 以 及 其 它 娱 乐 施 施

J|そうなんです、

E|I see,

C|知 道 了

J|それからプールや、サウナっていうのは、これも利用できるんですか。

E|and what about the swimming pool and sauna, [ah] can we use those too?

C|还 有 ， 有 没 有 游 泳 池 和 桑 纳

E|Only the outdoor pool is available right now

J|はい、御利用のほうは屋外プールのみになりますけれども、

C|现 在 只 有 户 外 泳 池 是 可 用 的

E|that's twenty dollars per person.

J|料金の方はお一人様、二十ドルになっております。

C|费用是每位二十美元

J|分かりました。|じゃ万一そこが満員の場合なんですが、ほかに、運動できる施設ってあります。

E|All right, what if the swimming pool is too crowded?|Are there any other facilities for exercising?

C|好的，假如游泳池太拥挤了还有其它的锻炼场所吗？

E|[well] You can use our gym for fifteen dollars

J|その場合は、ジムのほうが十五ドル、

C|您可以花十五美元使用我们的健身房

E|or the miniature golf course for twenty-five dollars.

J|そしてミニチュアゴルフコースが二十五ドルで御利用いただけます

C|或者花二十五美元租用小型高尔夫球场

E|We also have a sports deck and a jogging course at no charge.

J|スポーツデッキとジョギングコースのほうもございますけれども、こちらのほうは、無料で御利用いただけます。

C|我们也有一个运动甲板和一个跑步场可以免费使用

E|But we do not rent any equipment when the ship is in port.

J|ですけれども、停泊中の間<aida>に限りまして、[あ]スポーツ用品の貸し出しは行なっておりません。

C|但是当客船停泊时，我们不出租任何设备

J|分かりました [あー] それから、映画が見たいんですが、映画の上映ってあります。

E|I understand [ah] and I'd also like to see a movie there.|[ah] Do you have them?

C|知道了，我也喜欢看电影，船上有吗？

E|I'm sorry but our movie theater is closed

J|あいにくでございますけれども、ちょうど映画用の劇場の方はその(く)期間は閉まってるんです。

C|对不起，我们的影院已经关闭

E|but videos are shown in our theater bar.

J|そのかわりにシアターバーの方で、ビデオをお楽しみいただけます。

C|但是，您可以到录像厅去看录像

J|[あ]なるほどね、で、買物をしたいんですがねえ、

E|[oh] I see.|[ah] And I'd also like to do some shopping

C|知道了，还有，我也喜欢逛商店

J|広いショッピングエリアがあると聞いたんですが、そこ使えるんですか。

E|I heard that there's a large shopping area there.|[ah] Can we do some shopping there?

C|听说那儿有许多商店，我们能在那儿购物吗？

E|Of course, anytime.

J|はい、もちろんいつでも御利用願えます。

C|当然，随时欢迎

J|分かりました。|[えーと] それじゃあ [え] ツインの料金っていうのをいくらぐらいか教えてください。

E|All right, I understand.|[um] Could you tell me how much [ah] your twin rooms cost?

C|好的，我知道了，能告诉我你们的套房的价格吗？

E|[ah] Our twin rooms start at two hundred and fifty dollars per night.

J|ツインの料金は一泊二百五十ドル、から、でございます。

C|我们的套房的起价是每晚二百五十美元

J|二百五十ドルですね。

E|Two fifty.|All right.

C|二百五十，好的

J|じゃ買物は別にして、

E|Excluding what we might spend on our shopping,

C|除去 我们 可能 在 购物 上 的 花销

J|家内と二人で泊りますと、最低どれぐらいの予算でいろいろ楽しめるんでしょうかねえ。

E|if I stay there with my wife, [ah] how much should we expect to spend?

C|假如 我 和 我 妻子 一起 去 住宿 的话，大致 需要 花费 多少 钱

E|Including meals, you probably need about five hundred dollars.

J|お食事も含めまして、[まあ]五百ドルもあれば十分じゃないでしょうか。

C|包括 吃饭，你们 可能 需要 大约 五百 美元

E|We hope you can enjoy the voyage experience.

J|船旅気分を味わっていただけたと思います。

C|我们 希望 你们 能 享受 到 航海 的 经历

J|そうですか、じゃ[あの]分かりました、ありがとう。

E|All right.|[ah] Thank you for your information.

C|好的，谢谢 你 提供 信息

J|家内と相談してもう一度また電話します。

E|I'd like to talk about this with my wife and I'll get back to you later.

C|我 要 同 我 妻子 讨论 一下，然后 给你 答复

E|Okay, we'll be looking forward to receiving your call.

J|はい、お電話お待ち申し上げて

C|好的，我们 恭候 您 的 电话