TR-IT-0206

Baseline Acoustic Models for the Spoken Language Database (SDB/SLDB)

シンガー・ハラルド
Harald Singer

外村 政啓
Masahiro Tonomura

霍 強
Qiang Huo

石井 純
Jun Ishii

深田 俊明
Toshiaki Fukada

シュスター・マイク
Michael Schuster

1997.03

In this document we describe the creation of speaker-independent and gender-dependent acoustic models from ATR's speech and language (SLDB) and speech databases (SDB) both for phoneme context-independent and context-dependent case. Models and parameter files have been installed in an accessible file system inside ATR ITL. The software used was ATRSPREC version r04a10. This report can be accessed online via

- http://www.itl.atr.co.jp/~singer/publication/TR-IT-0206/baseline.ps.gz
- http://www.itl.atr.co.jp/~singer/publication/TR-IT-0206/baseline/baseline.html

目次

図目次

表目次

1 Introduction

To facilitate comparative experiments and for development of a speech-to-speech translation system for non-read (spontaneous) speech, the acoustic subgroup of department 1 decided to build acoustic baseline models and publish details about how they were created. This report gives the details and the baseline recognition results. This report does not intend to explain any of the underlying algorithms. Please consult the references.

We ran 2 sets of experiments: one in December 1996 with 18 test speakers and 280 training speakers (called Dec96) and the other one with 42 test speakers and 230 training speakers (called Jan97). As in the Dec96 experiments the number of test speakers was not sufficient to draw conclusions, we had to increase this number to a total of 42 speakers to increase the reliability of our results. The new speakers were chosen from the original training sets, i.e. the number of training speakers decreased. The Dec96 experiments are merely given for comparison in Appendix 付録 H .

In Section 2 we give details and references for the training and test sets of the database. In Section 3 we outline the procedure of creating acoustic models and give details about preprocessing. In Section 4 we give the experimental results and recommend acoustic models for comparative evaluations.

2 Training and Test Sets

Details of all training and test sets, e.g. mapping from conversation id to speaker id, are given in Appendix 付録 A .

We repartitioned the training and test set only using data from the SDB travel task to get a "pure" experiment. The testsets used were S1 and S2, which are both dealing with room reservations. S1 (117 utterances) and S2 (111 utterances) are both from the mono-lingual SDB database. Additionally, we choose a new testset, S4 (323 utterances), from the same database. We thus used 24 female and 18 male speakers with a total of 551 turns for recognition. This left us with a total of 99 male and 131 female speakers (see Table 1) for training.

表 1: Details of training data sets (Jan97). Numbers in brackets are for phone segments longer or equal to 40 msec. Frameshift is 10 msec.

	gender	#speakers	#pause units	#phones	#frames
M099	male	99	2348	53655 (37527)	343135 (300950)
F131	female	131	3164	74385 (54998)	504509 (453504)
A230	both	230	5512	128040 (92525)	847644 (754454)

Phone pair counts for male and female training sets were calculated from the transcription files and are given in Table 2 and Table 3. These tables show some of the difficulties for these training data, e.g. there is not even one occurence of a syllabic /ng/ followed by the vowel /u/.

Another problem is that roughly 25 % of the phonemes can not be used for topology training because they would be too short, i.e. less than 4 frames. On the other hand, especially these short phonemes might be highly coarticulated. With a 5 msec frameshift and a minimal length of 4 frames, the used number of phonemes would become 115655 and the number of frames 1625061.

表 2: Phone pair count in training set M099: e.g. 34 occurrences for /i/ followed by /a/

	a	i	u	e	o	ng	j	m	n	b	d	g	p	t	k	s	sh	ts	ch	z	zh	w	r	h	q
a	103	1012	8	60	72	210	62	89	268	32	184	45		121	376	686	188	39	59	27	18	48	353	52	58
i	34	178	1	18	51	277	26	646	202	24	147	96	13	552	255	103	341	59	77	14	68	48	124	44	34
u	12	84	171	11	152	76	36	39	230	20	139	176	2	100	351	77	60	26	6	7	51	17	182	34	18
e	19	95		169	127	121	84	98	74	36	275	351		100	113	576	132	17	11	8	23	61	179	41	81
o	37	109	20	21	804	64	122	275	513	28	307	83	1	127	361	64	288	26	67	37	32	48	376	160	50
ng	2	6		5	26	1	9	23	46	16	315	68	11	36	44	23	9	1	10	6	25	24	10	11	1
j	215		170		507																				
m	785	50	21	24	319		2																		
n	278	262	2	400	432		2																		
b	49	28	53	26	5		5																		
d	156	3		997	282																				
g	604	5	65	8	156		1																		
p	16		17	2	2		9																		
t	512	2		268	405																				
k	558	152	464	253	183		28																		
s	168		1118	95	262																				
sh		870					186																		
ts			183																						
ch		170		19			58																		
z	49		29	22	3																				
zh		105		1			149																		
w	282																								
r	216	313	220	264	200		30																		
h	279	53	62	60	112		11																		
q												4		18	117	83		15		5					

表 3: Phone pair count in training set F131: e.g. 54 occurrences for /i/ followed by /a/

	a	i	u	e	o	ng	j	m	n	b	d	g	p	t	k	s	sh	ts	ch	z	zh	w	r	h	q
a	160	1476	25	59	98	283	93	150	372	51	287	93		193	566	949	350	53	72	27	24	75	486	81	83
i	54	302	7	36	86	360	40	982	311	23	199	136	10	790	340	154	438	92	106	20	93	97	150	80	58
u	16	106	197	9	241	95	73	56	269	19	175	238	3	130	487	102	73	44	14	18	51	20	243	37	27
e	23	135	2	165	187	164	109	130	108	51	342	544		127	175	766	165	23	11	10	15	103	283	37	96
o	47	148	25	13	1024	81	149	409	788	42	396	109	3	192	455	110	403	53	51	66	39	68	453	208	58
ng	8	12		7	27	1	5	29	80	21	441	86	20	28	49	21	12	1	11	14	33	24	20	15	2
j	308		189		627																				
m	1185	85	46	45	444																				
n	379	400	1	593	620		5																		
b	72	40	63	31	6		7																		
d	196	10		1385	357																				
g	894	13	96	11	231		2																		
p	23	2	25		3		11																		
t	793	2		360	532																				
k	764	252	578	368	237		36																		
s	236		1597	114	324																				
sh		1258					244																		
ts			286																						
ch		201		25			69																		
z	60		49	37	15																				
zh		140					163																		
w	423																								
r	307	338	311	402	283		21																		
h	482	76	70	75	143		10																		
q												2		25	168	97	2	28	1	1					

3 Processing Flow

The general procedure of creating a context-dependent HMnet follows Figure 1[1]. Preprocessing conditions are described in Table 4.

表 4: Preprocessing Details

sampling rate	12kHz (downsampled from 16kHz)
preemphasis	0.98
sampling precision	16bit
frame shift	10 msec
frame length	20 msec
window type	Hamming
acoustic parameters	34 dimensional vector consisting of 16 order LPC Cepstrum, log power, 16 order ΔLPC Cepstrum, Δlog power
Δ window	triangular 100 msec (9 frames)

3.1 Viterbi Alignment of Transcribed Data

ML-SSS topology training currently needs phoneme boundary information. Time information for the SDB and SLDB databases however is only given for each pause unit (see Appendix 付録 G for an example of a .TRS file). As a first stage for training context-dependent models we therefore have to perform a Viterbi alignment with the best available acoustic model to get start and end time for each phoneme.

For each conversation side the segmentation information for each pause unit was used and a previously trained gender-dependent HMnet was adapted with this particular conversation side. No additional silences were added to start and end of each pause unit, i.e. the margin in Figure 2 for case a) was set to 0 msec. We then performed Viterbi alignment using both, label sequence and time information from the transcription files.

In preliminary experiments, we obtained better alignments with 5 msec frameshift than 10 msec frameshift so we performed all Viterbi alignments with 5 msec frameshift.

The previously trained gender-dependent HMnets were

```
/dept1/work10/basesystem/MODEL/HMnet_filled+sil.400.MHT
/dept1/work10/basesystem/MODEL/HMnet_filled+sil.400.FTK
```

where each HMnet consists of 400 states with 1 mixture and a 1 state, 10 mixture silence model. These models had been trained with the 2620 even-numbered words from the 5240 word Aset using SSS, separately for male speaker MHT and female speaker FTK.

Adaptation was performed with

```
Exe.adapt_HMnet -kn 6 -sr 0.0 -tm 3
```

i.e. only means and variances were retrained without any additional smoothing. Variances are only allowed to increase. This was the reason why we did not use `Exe.retrain_HMnet`

Viterbi alignment was performed with

```
Exe.viterbi_HMnet
```

Details and various comparisons can be found in[2].

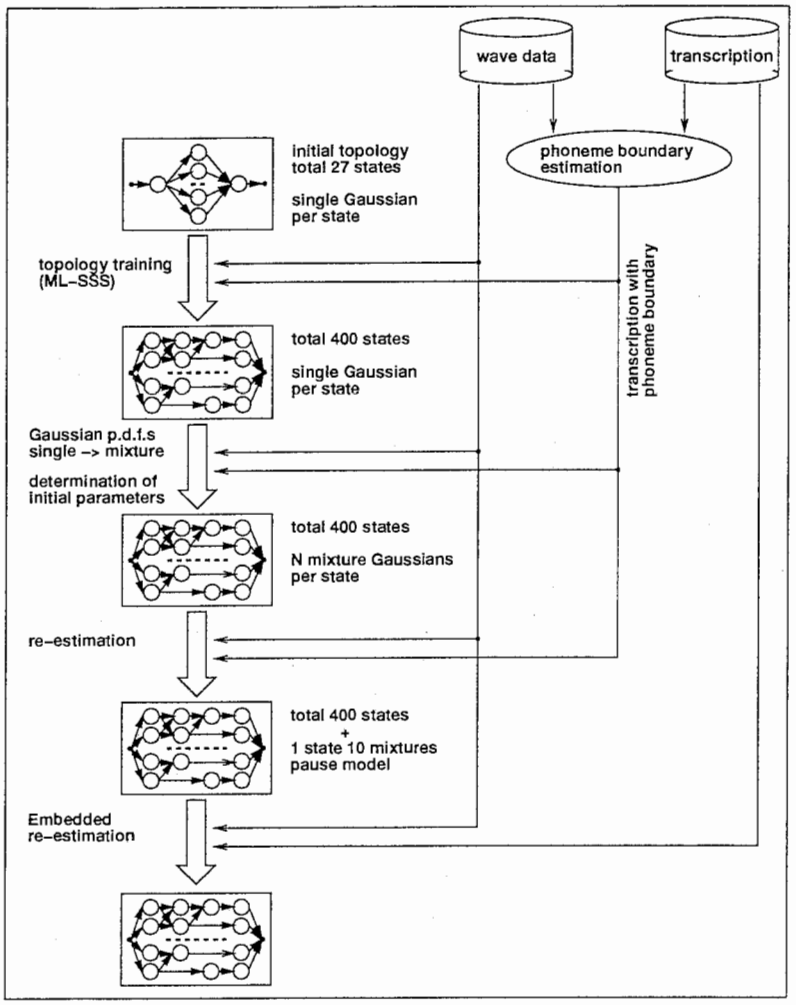

図 1: Flow for creating a context-dependendent, 400 state HMnet

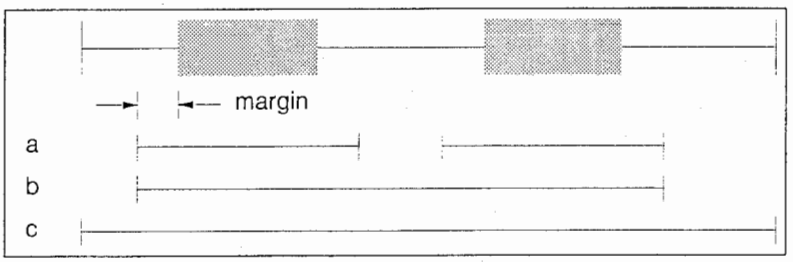

図 2: Use of segmentation information: a) pause unit with additional margin, b) from start of first pause unit until end of last pause unit, c) ignoring segmentation information

3.2 Training

(3.2.1) Topology Training

We used the ML-SSS algorithm[3] to generate context-dependent acoustic models. First, we trained gender-dependent topologies up to 1000 states. The initial topology used 27 states as shown in Figure 3 and the maximal path length was set to 4 states. The restriction to 4 states seems necessary as there are many phonemes in non-read speech that are quite short.

About 25% of the auto-aligned phonemes could not be used as they were shorter than 40 msec and they would not have fit the paths in the HMnet as we do not allow skips. Also note that in the Viterbi alignment step we have been using 5 msec frameshift, i.e. the shortest phoneme could be around 15 msec. [1]

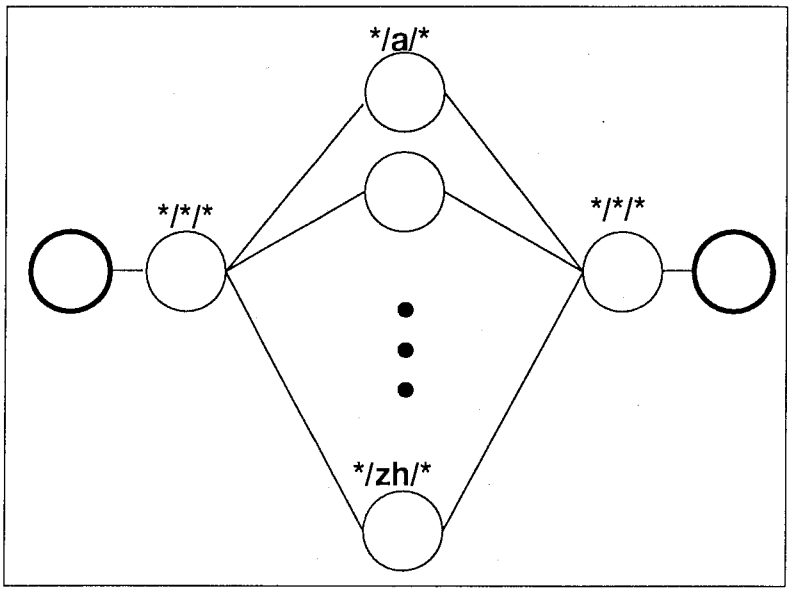

図 3: Initial model for context-dependent topology training

For the context-independent models, we created an HMnet structure as shown in Figure 4.

(3.2.2) Label-Based Retraining with Mixtures

We assume that everything that was not labeled as an utterance is non-speech. Correspondingly we trained speaker-independent and gender-dependent (context-independent) silence models. At first glance, gender-dependency does not seem to make much sense, but due to the preprocessing window, adjacent speech also might have an influence on the parameters. We trained 1 state, 10 mixture models for non-speech.

The state-sharing HMnet topologies were then retrained using standard Baum-Welch, using again the phoneme aligned data with 5 mixtures for each state. Initialization for the mixtures is done by a VQ[2]. This may be implemented using ATRSPREC with a command as follows:

```
Exe.retrain_HMnet -cm "label based, 10ms, no pause" -rp 1 -tt 0 -it 20
```

The separately trained 1 state, 10 mixture silence model was then concatenated with the label-retrained model as initial model for the embedded reestimation.

[1] In future research, this problem has to be addressed as coarticulation is especially strong for the short phoneme segments.

[2] Contrary to HTK, where the number of mixtures is gradually increased, we directly train the final number of desired mixtures per state.

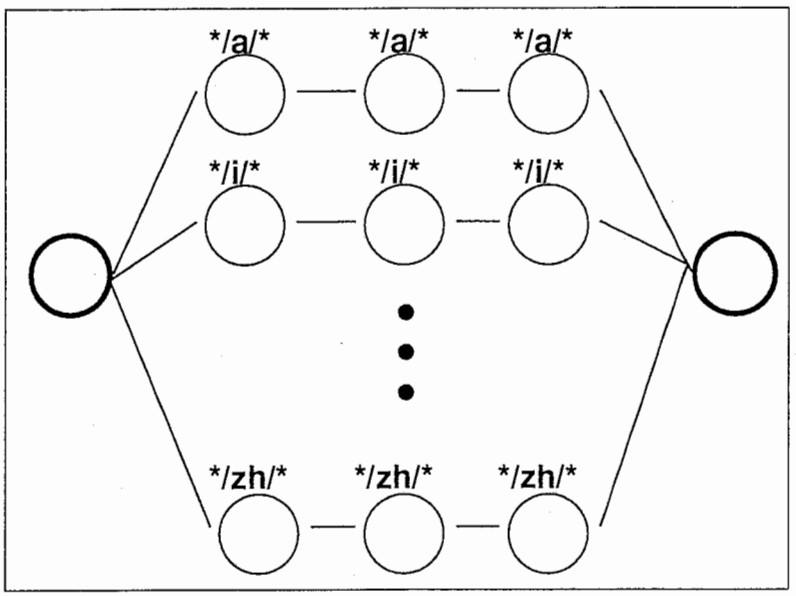

図 4: Context-independent topology

(3.2.3) Embedded Reestimation

Finally we did an embedded training run adding 30 msec to start and end of each pause unit, i.e. the margin in Figure 2 for case b) was set to 30 msec. For the labels, we concatenated the silence symbol label at the start and end of each pause unit. To speed up the retraining, we only did 10 iterations and used Viterbi training instead of Baum-Welch by using the following command:

```
Exe.retrain_HMnet -cm "embedded viterbi, 10ms, 30 ms pause" -rp 0 -tt 1 -it 10
```

Please see Appendix 付録 F for example logfiles of both, label based retraining and embedded reestimation.

3.3 Recognition

(3.3.1) Evaluation for Phoneme Recognition

Phonetic typewriter recognition experiments were done using ATRlattice with the phonotactic constraints of Japanese. Input was the complete wave file, i.e. the segmentation information in the .TRS files was not used (see Appendix 付録 G for an example). The multi-pass search of the lattice parser is described in [4, 5].

The phonotactic constraints of Japanese can be approximately expressed as phoneme pair grammar[6] which is displayed in a table, where each entry indicates which phoneme can follow which phoneme (see Table 5). We use 26 phoneme symbols and an additional $ symbol that stands for the start and end symbol.[3]. For example, /sh/ can only be followed by /i/ or /j/.

The recognition result is then DP-matched with the correct answer using equal penalty for insertion, deletion and substitution. Any /f/ phoneme in the transcription files was mapped to /h/. It is impossible to train /f/ models, because there are too few occurrencies in the training sets.

(3.3.2) Evaluation for Word Recognition

For a subset of acoustic models, i.e. the recommended models, (and a subset of the testset) we also run the lattice parser using variable-order class N-gram [7] as language model to perform word recognition. We report the first best accuracy and network accuracy.

[3]The $ symbol is actually implemented as symbol /5/ for start, which in the lexicon is mapped to a silence model and symbol /6/ which is a dummy model, i.e. not associated with any acoustic model

表 5: Phonotactic constraints of Japanese expressed as a phoneme pair grammar. Note: this is different from a bigram!

	$	-	a	i	u	e	o	k	s	sh	t	ch	ts	h	p	n	m	r	w	g	z	zh	d	b	j	q	ng
$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
-	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
i	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
u	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
o	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
k	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
s	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sh	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
t	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ch	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
ts	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
p	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
n	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
m	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
r	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
w	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
z	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
zh	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
b	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
j	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
q	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ng	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0

4 Experimental Results

Phoneme accuracy results for testsets S1, S2 and S4 using differently trained models are given in Tables 6 and 7. For more detailed results on a per-speaker basis please refer to Appendix 付録 D .

Currently ITL's recognition engine, ATRlattice, can use multiple sets of acoustic models in parallel but only if they have the same topology. The reason for this is that a unique lexical tree has to be constructed, where the nodes are linked to states in the HMnet. To use gender dependent models in parallel it is therefore necessary to train different sets of models from different datasets but with a common topology. Reestimation of the parameters (see Sections (3.2.2) and (3.2.3)) is thus performed with different data from those for topology training.

For example, the column labelled A230/M099 refers to the acoustic model where all 230 speakers have been used for topology training and retraining has been performed with the 99 male speaker training set.

表 6: Phoneme accuracy for male test speakers (Jan97): "CI X" stands for context independent, X mixtures left-to-right, 3 states HMM's. "CD X" stands for context dependent, X states, 5 mixture HMnet.

type	#gaussians	A230/A230	M099/M099	A230/M099	F131/M099
CI 15	1135	48.44	51.99	NA	NA
CD 400	2010	62.02	62.07	65.41	-
CD 600	3010	64.96	65.56	67.08	-
CD 800	4010	**66.35**	65.15	**69.02**	64.38
CD1000	5010	66.99	66.83	67.96	-

表 7: Phoneme accuracy for female test speakers (Jan97): "CI X" stands for context independent, X mixtures left-to-right, 3 states HMM's. "CD X" stands for context dependent, X states, 5 mixture HMnet.

type	#gaussians	A230/A230	F131/F131	A230/F131	M099/F131
CI 15	1135	52.37	52.80	NA	NA
CD 400	2010	70.02	71.62	70.78	-
CD 600	3010	71.95	72.79	72.55	-
CD 800	4010	**73.55**	71.71	**74.50**	71.42
CD1000	5010	73.51	73.94	74.37	-

4.1 Discussion

Best results were obtained with A230/M099 and A230/F131 CD 800 model. Accordingly, these models and the A230/A230 CD 800 model are our recommended models and have been installed for public access (see Appendix 付録 C). The recognition accuracies for the recommended models appear in bold face in Tables 6 and 7.

To analyze the recognition accuracy in more detail we investigated the dependency of the result on the number of unseen triphones, i.e. how many triphones for the conversation did not appear in the training data. The results in Figure 5 do not show a correlation between recognition accuracy and number of unseen triphones.

We also investigated the performance dependency on speaking rate as measured by the average number of morae per second based on the transcription files `.TRS`. Again, the results in Figure 6 do not show a correlation.

For comparison, word recognition results for male and femal testsets are given in Table 8. For more detailed results on a per-speaker basis please refer to Appendix 付録 D .

表 8: Word recognition accuracy for male and female testsets

	A230/M099	A230/F131	A230/A230
male testset	59.0	-	61.6
female testset	-	71.3	73.0

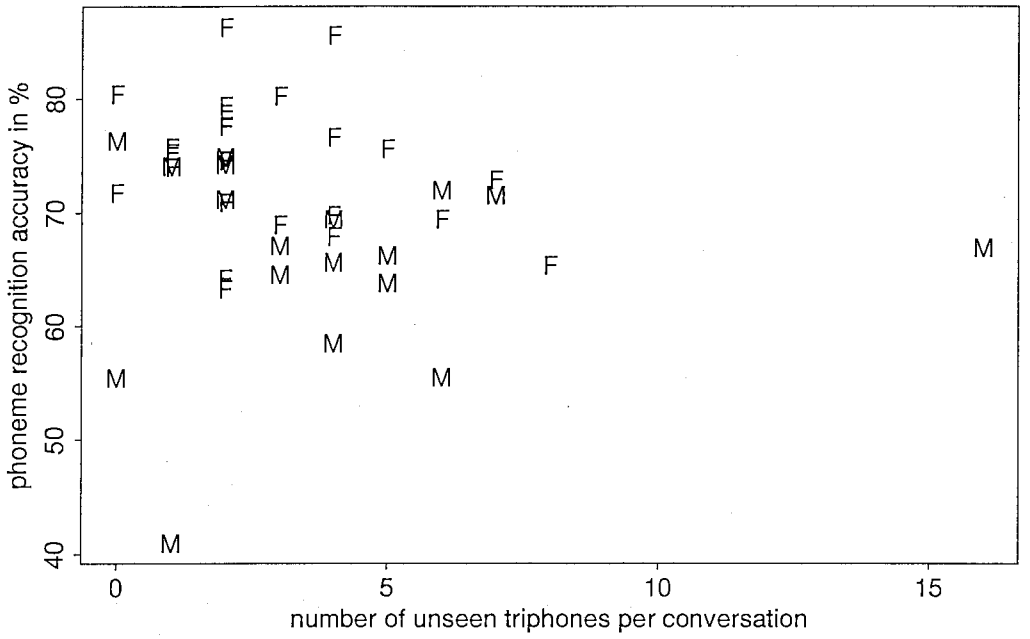

図 5: Recognition accuracy depending on number of unseen triphones in each test conversation side. M (male), F (female)

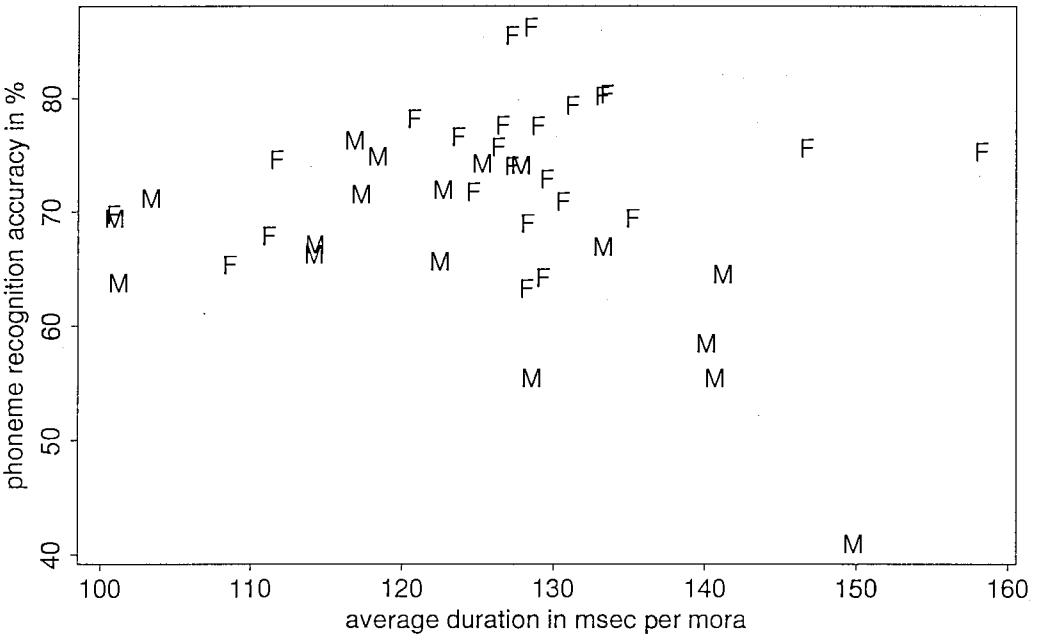

図 6: Recognition accuracy depending on speaking rate for each test conversation side. M (male), F (female)

5 Summary

In this report we investigated the creation of precise and robust acoustic models on the travel task subset of ATR's SDB database. The recommended models for comparative results are context-dependent 800 state models where topology training is performed using all training data (i.e. gender-independent). The HMnet with this topology is then retrained with gender-dependent data. Additionally, we also recommend a completely gender-independent model.

Due to time constraints, we did not test the robustness of the acoustic models with cross-task experiments.

Acknowledgments

The authors would like to thank Dr. Yamazaki, President, ATR Interpreting Telecommunications Research Laboratories, and Dr. Sagisaka, head of Department 1, for their continuous support of this work. We are also grateful to T. Shimizu and H. Yamamoto for developing the recognition engine, to H. Masataki for providing us with a language model and the other researchers at ATR ITL for discussions and comments.

参考文献

[1] M. Tonomura, T. Kosaka, and S. Matsunaga. Speaker adaptation based on transfer vector field smoothing using maximum a posteriori probability estimation. *Computer Speech and Language*, 10:117–132, 1996.

[2] M. Tonomura. Speaker-independent phone modeling using speaker specific phoneme alignment. In *Third Joint Meeting*, pages 987–992, Honolulu, 1996.

[3] M. Ostendorf and H. Singer. HMM topology design using maximum likelihood successive state splitting. *Computer Speech and Language*, 1997. (to appear).

[4] T. Shimizu, H. Yamamoto, S. Matsunaga, and Y. Sagisaka. Spontaneous dialogue speech recognition using cross-word context constrained word graphs. In *Proc. ICASSP*, pages 145–148, 1995.

[5] T. Shimizu, H. Singer, and Y. Sagisaka. Fast word-graph generation for spontaneous conversational speech translation. In *Proc. ICASSP*, pages –, Munich, 1997. (accepted).

[6] H. Singer and J. Takami. Speech recognition without grammar or vocabulary constraints. In *Proc. ICSLP*, pages 2207–2210, Yokohama, 1994.

[7] H. Masataki and Y. Sagisaka. Variable-order n-gram generation by word-class splitting and consecutive word grouping. In *Proc. ICASSP*, pages 188–191, Atlanta, 1996.

[8] A. Nakamura, S. Matsunaga, T. Shimizu, M. Tonomura, and Y. Sagisaka. Japanese speech database for robust speech recognition. In *Proc. ICSLP*, pages 2199–2192, Philadelphia, 1996.

付録 A Training and Test Sets

All information about training and test sets is recorded in files with extension `.ascii`. Here is an example line (index numbers added for readability):

```
  1       2      3    4     5    6     7  8          9              10    11   12     13
TAC70021 JTEXT   ENV  JMOR   WAV   TRS    X MMAHA      64db-18db      1994.00.00 16    A   CUSTOMER
```

The fields are

1. conversation id
2. does a Japanese orthographic transcription exist? (`JTEXT` or `X`)
3. does an environment file exist? (`ENV` or `X`)
4. does a Japanese morphological analysis file exist? (`JMOR` or `X`)
5. does a sampled wave file exist? (`WAV` or `X`)
6. does a phonetic transcription file exist? (`TRS` or `X`)
7. does a phonetic detailed label file exist? (`LBL` or `X`)
8. speaker id (first letter shows gender, i.e. `M` for male and `F` for female)
9. noise level
10. recording date
11. number of turns
12. conversation side (`A` for caller, `B` for callee, `C` for translator)
13. role

All wave and transcription files can then be accessed via

```
/DB/SDB/ALL/SPH/WAV/JAPANESE/$CONV/$CONV.$NUM.$SIDE.16k
/DB/SDB/ALL/SPH/TRS/JAPANESE/$CONV/$CONV.$NUM.$SIDE.TRS
```

where $SIDE is the conversation side (usually A or B) and $NUM is a 4 digit number, specifying a turn. This number is incremented in steps of 10 for the whole conversation.

More detailed information about test and training sets can be found at
http://www.itl.atr.co.jp/~singer/software/db/db/db.html.

`/DB/SDB/ALL/INFO/etc/testset/S1.ascii`

```
TAC70021 JTEXT   ENV  JMOR   WAV   TRS    X MMAHA      64db-18db      1994.00.00 16    A   CUSTOMER
TAC70019 JTEXT   ENV  JMOR   WAV   TRS    X FTOAR      64db-48db      1994.00.00 15    A   CUSTOMER
TAC70101 JTEXT   ENV  JMOR   WAV   TRS    X FYOOO      64db-24db      1994.00.00 20    A   CUSTOMER
TAC70103 JTEXT   ENV  JMOR   WAV   TRS    X MMAUC      64db-32db      1994.00.00 10    A   CUSTOMER
TAC70202 JTEXT   ENV  JMOR   WAV   TRS    X MSAHA      64db-32db      1994.00.00 11    A   CUSTOMER
TAC70203 JTEXT   ENV  JMOR   WAV   TRS    X FMASZ      64db-24db      1994.00.00 18    A   CUSTOMER
TAC70303 JTEXT   ENV  JMOR   WAV   TRS    X FYUKI      64db-32db      1994.00.00 19    A   CUSTOMER
TAC70304 JTEXT   ENV  JMOR   WAV   TRS    X MKEWA      64db-48db      1994.00.00  8    A   CUSTOMER
```

`/DB/SDB/ALL/INFO/etc/testset/SL2.ascii`

```
TAS12007 JTEXT   ENV  JMOR   WAV   TRS   LBL FYUYO       64db-16db      1994.05.09  6     A   CUSTOMER
TAS12010 JTEXT   ENV  JMOR   WAV   TRS   LBL FYOMA       56db-08db      1994.05.11  8     A   CUSTOMER
TAS12013 JTEXT   ENV  JMOR   WAV   TRS     X FKAKI       64db-32db      1994.07.26 14     A   CUSTOMER
TAS12026 JTEXT   ENV  JMOR   WAV   TRS     X MMAIS       64db-08db      1994.09.10 13     A   CUSTOMER
TAS22021 JTEXT   ENV  JMOR   WAV   TRS     X FHINI       48db-08db      1994.07.19 11     A   CUSTOMER
TAS22032 JTEXT   ENV  JMOR   WAV   TRS     X FJUMU       64db-32db      1994.06.20 12     A   CUSTOMER
TAS32007 JTEXT   ENV  JMOR   WAV   TRS   LBL MMAKO       64db-16db      1994.05.24 12     A   CUSTOMER
TAS32009 JTEXT   ENV  JMOR   WAV   TRS   LBL MNOSA       64db-24db      1994.06.07 12     A   CUSTOMER
TAS32015 JTEXT   ENV  JMOR   WAV   TRS     X MMASH       64db-32db      1994.07.16 11     A   CUSTOMER
TAS32016 JTEXT   ENV  JMOR   WAV   TRS     X MRYTA       64db-24db      1994.07.18  8     A   CUSTOMER
```

/DB/SDB/ALL/INFO/etc/testset/S2.ascii

```
TAC70015 JTEXT   ENV  JMOR   WAV   TRS     X FTAAO       64db-48db      1994.00.00 14     A   CUSTOMER
TAC70016 JTEXT   ENV  JMOR   WAV   TRS     X MNACH       64db-40db      1994.00.00 12     A   CUSTOMER
TAC70017 JTEXT   ENV  JMOR   WAV   TRS     X MMINA       72db-40db      1994.00.00 11     A   CUSTOMER
TAC70022 JTEXT   ENV  JMOR   WAV   TRS     X MJUHO       64db-40db      1994.00.00 13     A   CUSTOMER
TAC70023 JTEXT   ENV  JMOR   WAV   TRS     X MHITA       64db-40db      1994.00.00 17     A   CUSTOMER
TAC70102 JTEXT   ENV  JMOR   WAV   TRS     X FYOAZ       64db-40db      1994.00.00 18     A   CUSTOMER
TAC70201 JTEXT   ENV  JMOR   WAV   TRS     X FKOTS       64db-40db      1994.00.00 15     A   CUSTOMER
TAC70301 JTEXT   ENV  JMOR   WAV   TRS     X FTOHO       64db-40db      1994.00.00 11     A   CUSTOMER
```

/DB/SDB/ALL/INFO/etc/testset/S4.ascii

```
TCC70109 JTEXT ENV JMOR WAV TRS X MYUYO 72db-48db 1994.00.00 13 A CUSTOMER
TCC70103 JTEXT ENV JMOR WAV TRS X FRIYU 72db-48db 1994.00.00 9 A CUSTOMER
TCC71035 JTEXT ENV JMOR WAV TRS X MRYAR 66db-27db 1995.__.__ 11 A CUSTOMER
TCC71008 JTEXT ENV JMOR WAV TRS X MMAOX 61db-36db 1995.__.__ 17 A CUSTOMER
TCC71016 JTEXT ENV JMOR WAV TRS X FASSA 61db-38db 1995.__.__ 16 A CUSTOMER
TCC71001 JTEXT ENV JMOR WAV TRS X FASHI 58db-30db 1995.__.__ 12 B CLERK
TCS70070 JTEXT ENV JMOR WAV TRS X METYA 64db-32db 1994.00.00 8 A CUSTOMER
TCC70201 JTEXT ENV JMOR WAV TRS X MSHSZ 64db-40db 1994.00.00 7 A CUSTOMER
TCC70307 JTEXT ENV JMOR WAV TRS X MMAOK 64db-32db 1994.00.00 11 A CUSTOMER
TCS70055 JTEXT ENV JMOR WAV TRS X MMAMU 64db-32db 1994.00.00 11 A CUSTOMER
TCS70074 JTEXT ENV JMOR WAV TRS X MKENA 72db-40db 1994.00.00 12 A CUSTOMER
TCC70212 JTEXT ENV JMOR WAV TRS X MKEOO 64db-32db 1994.00.00 18 A CUSTOMER
TCS70034 JTEXT ENV JMOR WAV TRS X MKEWZ 64db-40db 1994.00.00 18 A CUSTOMER
TCC71007 JTEXT ENV JMOR WAV TRS X FJUSA 56db-40db 1995.__.__ 26 A CUSTOMER
TCS70013 JTEXT ENV JMOR WAV TRS X FYUNA 72db-40db 1994.00.00 7 A CUSTOMER
TSC71005 JTEXT ENV JMOR WAV TRS X FAKMX 62db-33db 1995.__.__ 15 B CLERK
TCS70023 JTEXT ENV JMOR WAV TRS X FKYYA 64db-40db 1994.00.00 14 A CUSTOMER
TCS70059 JTEXT ENV JMOR WAV TRS X FSAWA 64db-40db 1994.00.00 6 A CUSTOMER
TCS70082 JTEXT ENV JMOR WAV TRS X FNOTA 72db-48db 1994.00.00 8 A CUSTOMER
TCS70004 JTEXT ENV JMOR WAV TRS X FKENA 48db-16db 1994.00.00 15 B CLERK
TCS70010 JTEXT ENV JMOR WAV TRS X FAYYA 64db-40db 1994.00.00 10 A CUSTOMER
TCS70028 JTEXT ENV JMOR WAV TRS X FKITA 64db-40db 1994.00.00 13 A CUSTOMER
TCS70025 JTEXT ENV JMOR WAV TRS X FYOSU 64db-48db 1994.00.00 9 A CUSTOMER
TSC71013 JTEXT ENV JMOR WAV TRS X FHAOK 65db-33db 1995.__.__ 14 A CUSTOMER
TCS70020 JTEXT ENV JMOR WAV TRS X FYUNI 64db-24db 1994.00.00 11 A CUSTOMER
TCS70047 JTEXT ENV JMOR WAV TRS X FYUKO 64db-48db 1994.00.00 12 A CUSTOMER
```

/DB/SDB/ALL/INFO/etc/trainingset/T_M_0099.ascii

```
TCC70110 JTEXT ENV JMOR WAV TRS X MYUAI 64db-40db 1994.00.00 7 A CUSTOMER
TCC60200 JTEXT ENV JMOR WAV TRS X MAKYO 64db-24db 1994.11.22 11 A CUSTOMER
TCC70403 JTEXT ENV JMOR WAV TRS X MSHMA 64db-32db 1994.00.00 11 A CUSTOMER
TCC70108 JTEXT ENV JMOR WAV TRS X MNAMO 72db-48db 1994.00.00 15 A CUSTOMER
TCC70404 JTEXT ENV JMOR WAV TRS X MYANA 64db-32db 1994.00.00 12 A CUSTOMER
TCC71002 JTEXT ENV JMOR WAV TRS X MMAKW 63db-33db 1995.__.__ 9 A CUSTOMER
TCC70099 JTEXT ENV JMOR WAV TRS X MSHMI 64db-40db 1994.00.00 11 A CUSTOMER
TCC71038 JTEXT ENV JMOR WAV TRS X MMASY 61db-30db 1995.__.__ 10 A CUSTOMER
TCC71037 JTEXT ENV JMOR WAV TRS X MTAYZ 62db-27db 1995.__.__ 11 A CUSTOMER
TCC71017 JTEXT ENV JMOR WAV TRS X MMONA 58db-37db 1995.__.__ 16 A CUSTOMER
TCC70092 JTEXT ENV JMOR WAV TRS X MYOSO 64db-40db 1994.00.00 7 A CUSTOMER
TCC70501 JTEXT ENV JMOR WAV TRS X MTSMI 64db-32db 1994.00.00 10 A CUSTOMER
TCC70098 JTEXT ENV JMOR WAV TRS X MTEOO 64db-40db 1994.00.00 9 A CUSTOMER
TCC70506 JTEXT ENV JMOR WAV TRS X MMATZ 64db-24db 1994.00.00 10 A CUSTOMER
TCC70502 JTEXT ENV JMOR WAV TRS X MOSKA 72db-32db 1994.00.00 6 A CUSTOMER
TCC71015 JTEXT ENV JMOR WAV TRS X MJUIS 60db-35db 1995.__.__ 12 B CLERK
```

```
TSC71010 JTEXT ENV JMOR WAV TRS X MHIOY 65db-28db 1995.__.__ 14 A CUSTOMER
TCS70027 JTEXT ENV JMOR WAV TRS X MMATS 72db-40db 1994.00.00 14 A CUSTOMER
TCS70054 JTEXT ENV JMOR WAV TRS X MSANU 64db-32db 1994.00.00 12 A CUSTOMER
TCC70312 JTEXT ENV JMOR WAV TRS X MTADA 64db-32db 1994.00.00 9 A CUSTOMER
TCS70050 JTEXT ENV JMOR WAV TRS X MKAIS 64db-40db 1994.00.00 9 A CUSTOMER
TSC71001 JTEXT ENV JMOR WAV TRS X MKAYY 60db-26db 1995.__.__ 21 B CLERK
TCC70303 JTEXT ENV JMOR WAV TRS X MHIMO 64db-32db 1994.00.00 16 A CUSTOMER
TCS70029 JTEXT ENV JMOR WAV TRS X MTONI 64db-40db 1994.00.00 9 A CUSTOMER
TCS70086 JTEXT ENV JMOR WAV TRS X MKASE 64db-32db 1994.00.00 12 A CUSTOMER
TCS70062 JTEXT ENV JMOR WAV TRS X MRYHA 72db-40db 1994.00.00 10 A CUSTOMER
TCS70061 JTEXT ENV JMOR WAV TRS X MYUKU 64db-32db 1994.00.00 6 A CUSTOMER
TCS70026 JTEXT ENV JMOR WAV TRS X MKUHA 64db-48db 1994.00.00 10 A CUSTOMER
TCC70210 JTEXT ENV JMOR WAV TRS X MHISU 64db-24db 1994.00.00 12 A CUSTOMER
TCC71032 JTEXT ENV JMOR WAV TRS X MMAKV 62db-22db 1995.__.__ 17 A CUSTOMER
TCC70313 JTEXT ENV JMOR WAV TRS X MTAYA 64db-32db 1994.00.00 8 A CUSTOMER
TCC70205 JTEXT ENV JMOR WAV TRS X MYOIS 64db-32db 1994.00.00 13 A CUSTOMER
TCC70204 JTEXT ENV JMOR WAV TRS X MYUIM 64db-32db 1994.00.00 7 A CUSTOMER
TCS70003 JTEXT ENV JMOR WAV TRS X MMAFU 64db-24db 1994.00.00 9 A CUSTOMER
TSC71001 JTEXT ENV JMOR WAV TRS X MSATA 60db-32db 1995.__.__ 20 A CUSTOMER
TCC71033 JTEXT ENV JMOR WAV TRS X MMAUZ 59db-26db 1995.__.__ 21 A CUSTOMER
TCC70312 JTEXT ENV JMOR WAV TRS X MHIKZ 64db-32db 1994.00.00 10 B CLERK
TCS70060 JTEXT ENV JMOR WAV TRS X MYOKU 64db-48db 1994.00.00 10 A CUSTOMER
TCS70041 JTEXT ENV JMOR WAV TRS X MTAHI 64db-40db 1994.00.00 9 A CUSTOMER
TCC70206 JTEXT ENV JMOR WAV TRS X MTAOO 56db-24db 1994.00.00 13 A CUSTOMER
TCS70084 JTEXT ENV JMOR WAV TRS X MYUMI 64db-32db 1994.00.00 17 A CUSTOMER
TCC70311 JTEXT ENV JMOR WAV TRS X MTOAZ 64db-32db 1994.00.00 13 A CUSTOMER
TSC71011 JTEXT ENV JMOR WAV TRS X MYAME 63db-27db 1995.__.__ 22 A CUSTOMER
TSC71012 JTEXT ENV JMOR WAV TRS X MHISX 58db-35db 1995.__.__ 18 A CUSTOMER
TSC71008 JTEXT ENV JMOR WAV TRS X MMAMX 59db-28db 1995.__.__ 11 A CUSTOMER
TCS70076 JTEXT ENV JMOR WAV TRS X MHIUE 64db-32db 1994.00.00 6 A CUSTOMER
TCC71020 JTEXT ENV JMOR WAV TRS X MYUFU 60db-29db 1995.__.__ 17 A CUSTOMER
TCC70308 JTEXT ENV JMOR WAV TRS X MHIYO 64db-24db 1994.00.00 13 A CUSTOMER
TCS70033 JTEXT ENV JMOR WAV TRS X MHIHI 64db-40db 1994.00.00 8 A CUSTOMER
TCS70015 JTEXT ENV JMOR WAV TRS X MMAAZ 64db-40db 1994.00.00 7 A CUSTOMER
TCS70075 JTEXT ENV JMOR WAV TRS X MOSIG 64db-32db 1994.00.00 11 A CUSTOMER
TCC70202 JTEXT ENV JMOR WAV TRS X MHITO 72db-40db 1994.00.00 11 A CUSTOMER
TCS70037 JTEXT ENV JMOR WAV TRS X MHIMA 64db-40db 1994.00.00 13 A CUSTOMER
TSC71003 JTEXT ENV JMOR WAV TRS X MSHOO 56db-37db 1995.__.__ 23 A CUSTOMER
TCC71031 JTEXT ENV JMOR WAV TRS X MKENO 62db-24db 1995.__.__ 20 A CUSTOMER
TCS70056 JTEXT ENV JMOR WAV TRS X MTOKI 72db-40db 1994.00.00 10 A CUSTOMER
TCS70018 JTEXT ENV JMOR WAV TRS X MTAAI 64db-40db 1994.00.00 11 A CUSTOMER
TCC71006 JTEXT ENV JMOR WAV TRS X MMIKA 55db-35db 1995.__.__ 30 A CUSTOMER
TCC70310 JTEXT ENV JMOR WAV TRS X MKOTA 64db-32db 1994.00.00 11 A CUSTOMER
TCC70305 JTEXT ENV JMOR WAV TRS X MMOTA 64db-32db 1994.00.00 9 A CUSTOMER
TCC70208 JTEXT ENV JMOR WAV TRS X MDAMA 64db-40db 1994.00.00 12 A CUSTOMER
TCS70040 JTEXT ENV JMOR WAV TRS X MTOHO 64db-32db 1994.00.00 12 A CUSTOMER
TCS70063 JTEXT ENV JMOR WAV TRS X MTOKZ 64db-32db 1994.00.00 6 A CUSTOMER
TSC71009 JTEXT ENV JMOR WAV TRS X MTAIZ 62db-23db 1995.__.__ 13 A CUSTOMER
TCS70001 JTEXT ENV JMOR WAV TRS X MMAII 56db-24db 1994.00.00 7 A CUSTOMER
TCS70038 JTEXT ENV JMOR WAV TRS X MYOMA 64db-48db 1994.00.00 12 A CUSTOMER
TCS70042 JTEXT ENV JMOR WAV TRS X MCHSA 72db-32db 1994.00.00 9 A CUSTOMER
TCS70022 JTEXT ENV JMOR WAV TRS X MFUAK 56db-24db 1994.00.00 13 A CUSTOMER
TCS70014 JTEXT ENV JMOR WAV TRS X MMANA 72db-40db 1994.00.00 9 A CUSTOMER
TCS70080 JTEXT ENV JMOR WAV TRS X MTAWA 64db-32db 1994.00.00 9 A CUSTOMER
TCS70024 JTEXT ENV JMOR WAV TRS X MDAIN 64db-32db 1994.00.00 9 A CUSTOMER
TCS70002 JTEXT ENV JMOR WAV TRS X MTOKO 56db-24db 1994.00.00 7 A CUSTOMER
TCS70032 JTEXT ENV JMOR WAV TRS X MTASU 56db-40db 1994.00.00 9 A CUSTOMER
TCC71034 JTEXT ENV JMOR WAV TRS X MSHIS 59db-27db 1995.__.__ 20 A CUSTOMER
TAC60298 JTEXT ENV JMOR WAV TRS X MRYMO 56db-24db 1994.12.05 18 A CUSTOMER
TAC70027 JTEXT ENV JMOR WAV TRS X MNOOG 72db-40db 1994.00.00 9 B CLERK
TAC70028 JTEXT ENV JMOR WAV TRS X MTANA 64db-32db 1994.00.00 14 A CUSTOMER
TAC70029 JTEXT ENV JMOR WAV TRS X MHIAT 64db-32db 1994.00.00 20 A CUSTOMER
TAC70030 JTEXT ENV JMOR WAV TRS X MAKKI 64db-48db 1994.00.00 12 B CLERK
TAC71001 JTEXT ENV JMOR WAV TRS X MYUYA 57db-34db 1995.__.__ 30 A CUSTOMER
TAC71007 JTEXT ENV JMOR WAV TRS X MNOSH 60db-33db 1995.__.__ 18 A CUSTOMER
TAC71013 JTEXT ENV JMOR WAV TRS X MSASU 60db-27db 1995.__.__ 18 A CUSTOMER
TAS70003 JTEXT ENV JMOR WAV TRS X MTEIS 64db-32db 1994.00.00 16 A CUSTOMER
TAS70004 JTEXT ENV JMOR WAV TRS X MTAMO 64db-32db 1994.00.00 11 A CUSTOMER
TAS70009 JTEXT ENV JMOR WAV TRS X MTATA 64db-48db 1994.00.00 15 A CUSTOMER
TAS70010 JTEXT ENV JMOR WAV TRS X MSHSO 64db-40db 1994.00.00 15 A CUSTOMER
TAS70011 JTEXT ENV JMOR WAV TRS X MHAMO 72db-40db 1994.00.00 20 A CUSTOMER
TAS70013 JTEXT ENV JMOR WAV TRS X MKETA 64db-32db 1994.00.00 13 A CUSTOMER
TAC60186 JTEXT ENV JMOR WAV TRS X MHIKY 56db-32db 1994.11.21 7 A CUSTOMER
TAC60228 JTEXT ENV JMOR WAV TRS X MTOMA 64db-24db 1994.11.24 11 A CUSTOMER
TAC60241 JTEXT ENV JMOR WAV TRS X MOSMI 72db-24db 1994.12.01 9 A CUSTOMER
TAC60312 JTEXT ENV JMOR WAV TRS X MTATZ 56db-24db 1994.12.05 7 A CUSTOMER
TAC60367 JTEXT ENV JMOR WAV TRS X MTEMO 56db-24db 1994.12.06 8 A CUSTOMER
TAC71002 JTEXT ENV JMOR WAV TRS X MTEMI 57db-38db 1995.__.__ 15 A CUSTOMER
TAC71004 JTEXT ENV JMOR WAV TRS X MYAMZ 58db-40db 1995.__.__ 14 A CUSTOMER
TAC71008 JTEXT ENV JMOR WAV TRS X MRYSA 57db-38db 1995.__.__ 20 A CUSTOMER
TAC71009 JTEXT ENV JMOR WAV TRS X MKOTX 56db-35db 1995.__.__ 13 A CUSTOMER
TAC71010 JTEXT ENV JMOR WAV TRS X MTAKX 61db-21db 1995.__.__ 7 A CUSTOMER
TAC71015 JTEXT ENV JMOR WAV TRS X MHISW 57db-29db 1995.__.__ 11 A CUSTOMER
```

/DB/SDB/ALL/INFO/etc/trainingset/T_F_0131.ascii

```
TCC70102 JTEXT ENV JMOR WAV TRS X FKOTE 72db-32db 1994.00.00 9 A CUSTOMER
TCC70109 JTEXT ENV JMOR WAV TRS X FMIIS 72db-24db 1994.00.00 14 B CLERK
TCC71027 JTEXT ENV JMOR WAV TRS X FREII 65db-34db 1995.__.__ 5 A CUSTOMER
TCC70107 JTEXT ENV JMOR WAV TRS X FYUKZ 72db-48db 1994.00.00 7 A CUSTOMER
TCC70104 JTEXT ENV JMOR WAV TRS X FREOO 72db-48db 1994.00.00 13 A CUSTOMER
TCC70111 JTEXT ENV JMOR WAV TRS X FKIFU 64db-40db 1994.00.00 13 A CUSTOMER
TCC70402 JTEXT ENV JMOR WAV TRS X FYUAR 64db-48db 1994.00.00 20 A CUSTOMER
TCC70106 JTEXT ENV JMOR WAV TRS X FSHHA 72db-40db 1994.00.00 16 A CUSTOMER
TCC71012 JTEXT ENV JMOR WAV TRS X FAYHZ 60db-44db 1995.__.__ 7 A CUSTOMER
TCC70101 JTEXT ENV JMOR WAV TRS X FCHMA 64db-24db 1994.00.00 9 A CUSTOMER
TCC70097 JTEXT ENV JMOR WAV TRS X FAKMI 64db-48db 1994.00.00 13 A CUSTOMER
TCC70093 JTEXT ENV JMOR WAV TRS X FKEKA 64db-40db 1994.00.00 9 A CUSTOMER
TCC71036 JTEXT ENV JMOR WAV TRS X FKAAR 60db-28db 1995.__.__ 11 A CUSTOMER
TCC71029 JTEXT ENV JMOR WAV TRS X FNOHX 67db-26db 1995.__.__ 8 A CUSTOMER
TCC70507 JTEXT ENV JMOR WAV TRS X FSEUT 64db-32db 1994.00.00 13 A CUSTOMER
TCC71003 JTEXT ENV JMOR WAV TRS X FAKMU 53db-41db 1995.__.__ 16 A CUSTOMER
TCC71015 JTEXT ENV JMOR WAV TRS X FMIHX 57db-35db 1995.__.__ 11 A CUSTOMER
TCC71001 JTEXT ENV JMOR WAV TRS X FMAIM 52db-37db 1995.__.__ 11 A CUSTOMER
TCC70094 JTEXT ENV JMOR WAV TRS X FKEIM 64db-48db 1994.00.00 10 A CUSTOMER
TCC70505 JTEXT ENV JMOR WAV TRS X FHATA 64db-32db 1994.00.00 7 A CUSTOMER
TCC71028 JTEXT ENV JMOR WAV TRS X FJUSZ 62db-22db 1995.__.__ 11 A CUSTOMER
TCC70096 JTEXT ENV JMOR WAV TRS X FYUKA 64db-32db 1994.00.00 8 A CUSTOMER
TCC71004 JTEXT ENV JMOR WAV TRS X FTAKZ 61db-39db 1995.__.__ 11 A CUSTOMER
TCC70095 JTEXT ENV JMOR WAV TRS X FCHKI 64db-40db 1994.00.00 9 A CUSTOMER
TCC70504 JTEXT ENV JMOR WAV TRS X FCHSA 56db-32db 1994.00.00 10 A CUSTOMER
TCC70091 JTEXT ENV JMOR WAV TRS X FEIOO 72db-48db 1994.00.00 10 A CUSTOMER
TCC70090 JTEXT ENV JMOR WAV TRS X FTSOG 64db-40db 1994.00.00 8 A CUSTOMER
TCC71009 JTEXT ENV JMOR WAV TRS X FYUSZ 62db-33db 1995.__.__ 11 A CUSTOMER
TCC71030 JTEXT ENV JMOR WAV TRS X FMION 66db-25db 1995.__.__ 7 A CUSTOMER
TCC70207 JTEXT ENV JMOR WAV TRS X FKESZ 64db-48db 1994.00.00 9 A CUSTOMER
TCS70065 JTEXT ENV JMOR WAV TRS X FYOTZ 72db-48db 1994.00.00 20 A CUSTOMER
TCS70079 JTEXT ENV JMOR WAV TRS X FAKKA 64db-40db 1994.00.00 13 A CUSTOMER
TCC70304 JTEXT ENV JMOR WAV TRS X FCHHA 64db-40db 1994.00.00 9 A CUSTOMER
TCS70046 JTEXT ENV JMOR WAV TRS X FMITZ 72db-40db 1994.00.00 11 A CUSTOMER
TCC71021 JTEXT ENV JMOR WAV TRS X FMIOW 55db-42db 1995.__.__ 22 A CUSTOMER
TCS70073 JTEXT ENV JMOR WAV TRS X FMIYA 64db-40db 1994.00.00 8 A CUSTOMER
TCS70008 JTEXT ENV JMOR WAV TRS X FYUAM 64db-32db 1994.00.00 7 A CUSTOMER
TCS70089 JTEXT ENV JMOR WAV TRS X FEIAO 64db-40db 1994.00.00 13 A CUSTOMER
TCS70069 JTEXT ENV JMOR WAV TRS X FTANA 72db-48db 1994.00.00 12 A CUSTOMER
TCC71019 JTEXT ENV JMOR WAV TRS X FAKSU 56db-40db 1995.__.__ 21 A CUSTOMER
TCS70072 JTEXT ENV JMOR WAV TRS X FAKYA 64db-32db 1994.00.00 8 A CUSTOMER
TCC70203 JTEXT ENV JMOR WAV TRS X FKASA 64db-32db 1994.00.00 12 A CUSTOMER
TCS70064 JTEXT ENV JMOR WAV TRS X FTOOK 64db-48db 1994.00.00 10 A CUSTOMER
TCS70030 JTEXT ENV JMOR WAV TRS X FEIFU 64db-40db 1994.00.00 12 A CUSTOMER
TCC70306 JTEXT ENV JMOR WAV TRS X FAMIS 64db-40db 1994.00.00 10 A CUSTOMER
TCS70068 JTEXT ENV JMOR WAV TRS X FSUTS 64db-48db 1994.00.00 17 A CUSTOMER
TCS70039 JTEXT ENV JMOR WAV TRS X FEMNA 56db-32db 1994.00.00 9 A CUSTOMER
TCC71026 JTEXT ENV JMOR WAV TRS X FSEMI 63db-21db 1995.__.__ 23 A CUSTOMER
TCS70006 JTEXT ENV JMOR WAV TRS X FNOSA 64db-40db 1994.00.00 11 A CUSTOMER
TCC71014 JTEXT ENV JMOR WAV TRS X FAKYZ 58db-42db 1995.__.__ 19 A CUSTOMER
TCS70066 JTEXT ENV JMOR WAV TRS X FMIOO 64db-48db 1994.00.00 9 A CUSTOMER
TCC71005 JTEXT ENV JMOR WAV TRS X FASYA 58db-43db 1995.__.__ 33 A CUSTOMER
TCC70213 JTEXT ENV JMOR WAV TRS X FMAMO 64db-32db 1994.00.00 10 A CUSTOMER
TCS70085 JTEXT ENV JMOR WAV TRS X FSAKO 64db-24db 1994.00.00 17 A CUSTOMER
TCS70067 JTEXT ENV JMOR WAV TRS X FTOOO 72db-48db 1994.00.00 8 A CUSTOMER
TCS70077 JTEXT ENV JMOR WAV TRS X FYOTO 64db-40db 1994.00.00 10 A CUSTOMER
TCC70212 JTEXT ENV JMOR WAV TRS X FTAYO 64db-32db 1994.00.00 19 B CLERK
TCS70044 JTEXT ENV JMOR WAV TRS X FMITO 64db-48db 1994.00.00 9 A CUSTOMER
TCC70302 JTEXT ENV JMOR WAV TRS X FMAKY 64db-32db 1994.00.00 16 A CUSTOMER
TCS70019 JTEXT ENV JMOR WAV TRS X FEMNI 64db-40db 1994.00.00 14 A CUSTOMER
TCS70051 JTEXT ENV JMOR WAV TRS X FTOKO 64db-32db 1994.00.00 13 A CUSTOMER
TCS70088 JTEXT ENV JMOR WAV TRS X FRIMI 64db-48db 1994.00.00 15 A CUSTOMER
TCS70058 JTEXT ENV JMOR WAV TRS X FMOMA 72db-40db 1994.00.00 6 A CUSTOMER
TCS70021 JTEXT ENV JMOR WAV TRS X FMISH 72db-40db 1994.00.00 9 A CUSTOMER
TSC71007 JTEXT ENV JMOR WAV TRS X FNOSZ 65db-27db 1995.__.__ 16 A CUSTOMER
TCS70034 JTEXT ENV JMOR WAV TRS X FKEKO 64db-24db 1994.00.00 21 B CLERK
TCC71010 JTEXT ENV JMOR WAV TRS X FMINP 55db-42db 1995.__.__ 19 A CUSTOMER
TCS70009 JTEXT ENV JMOR WAV TRS X FRUKO 64db-40db 1994.00.00 7 A CUSTOMER
TCS70049 JTEXT ENV JMOR WAV TRS X FAYIS 64db-48db 1994.00.00 10 A CUSTOMER
TCC70309 JTEXT ENV JMOR WAV TRS X FSAFU 64db-32db 1994.00.00 14 A CUSTOMER
TCC70211 JTEXT ENV JMOR WAV TRS X FAKKI 64db-32db 1994.00.00 8 A CUSTOMER
TCS70016 JTEXT ENV JMOR WAV TRS X FHIMU 64db-40db 1994.00.00 16 A CUSTOMER
TCS70048 JTEXT ENV JMOR WAV TRS X FYUOO 72db-48db 1994.00.00 11 A CUSTOMER
TCS70053 JTEXT ENV JMOR WAV TRS X FKISU 64db-48db 1994.00.00 8 A CUSTOMER
TCC71040 JTEXT ENV JMOR WAV TRS X FSHTA 61db-31db 1995.__.__ 27 A CUSTOMER
TCS70081 JTEXT ENV JMOR WAV TRS X FYUTZ 64db-48db 1994.00.00 14 A CUSTOMER
TCS70078 JTEXT ENV JMOR WAV TRS X FTOTZ 64db-40db 1994.00.00 11 A CUSTOMER
```

```
TCC71039 JTEXT ENV JMOR WAV TRS X FAYKZ 62db-27db 1995.__.__ 29 A CUSTOMER
TCS70083 JTEXT ENV JMOR WAV TRS X FNAIR 64db-32db 1994.00.00 15 A CUSTOMER
TCC71013 JTEXT ENV JMOR WAV TRS X FSAOY 57db-43db 1995.__.__ 21 A CUSTOMER
TCS70035 JTEXT ENV JMOR WAV TRS X FCHYO 64db-40db 1994.00.00 7 A CUSTOMER
TCS70043 JTEXT ENV JMOR WAV TRS X FTATO 72db-40db 1994.00.00 9 A CUSTOMER
TCS70011 JTEXT ENV JMOR WAV TRS X FCHKA 64db-40db 1994.00.00 12 A CUSTOMER
TCS70007 JTEXT ENV JMOR WAV TRS X FYAYO 64db-32db 1994.00.00 9 A CUSTOMER
TCS70071 JTEXT ENV JMOR WAV TRS X FRESU 64db-48db 1994.00.00 5 A CUSTOMER
TCS70052 JTEXT ENV JMOR WAV TRS X FWAOZ 64db-48db 1994.00.00 9 A CUSTOMER
TCS70036 JTEXT ENV JMOR WAV TRS X FKOMI 56db-40db 1994.00.00 8 A CUSTOMER
TCS70087 JTEXT ENV JMOR WAV TRS X FJUSH 64db-40db 1994.00.00 10 A CUSTOMER
TCS70005 JTEXT ENV JMOR WAV TRS X FSASU 48db-24db 1994.00.00 8 A CUSTOMER
TCS70017 JTEXT ENV JMOR WAV TRS X FKESH 64db-40db 1994.00.00 11 A CUSTOMER
TCC70301 JTEXT ENV JMOR WAV TRS X FTATA 64db-40db 1994.00.00 14 A CUSTOMER
TCS70057 JTEXT ENV JMOR WAV TRS X FMITA 72db-48db 1994.00.00 8 A CUSTOMER
TCS70004 JTEXT ENV JMOR WAV TRS X FYOTA 48db-16db 1994.00.00 14 A CUSTOMER
TSC71006 JTEXT ENV JMOR WAV TRS X FHIOY 62db-25db 1995.__.__ 16 A CUSTOMER
TCS70012 JTEXT ENV JMOR WAV TRS X FKEHO 64db-40db 1994.00.00 8 A CUSTOMER
TCC71011 JTEXT ENV JMOR WAV TRS X FYUFY 60db-40db 1995.__.__ 16 A CUSTOMER
TCC71024 JTEXT ENV JMOR WAV TRS X FTESU 60db-35db 1995.__.__ 17 A CUSTOMER
TCS70045 JTEXT ENV JMOR WAV TRS X FCHYZ 64db-48db 1994.00.00 10 A CUSTOMER
TAC60172 JTEXT ENV JMOR WAV TRS X FYONA 64db-32db 1994.11.21 12 A CUSTOMER
TAC60214 JTEXT ENV JMOR WAV TRS X FNOMU 64db-32db 1994.11.24 18 A CUSTOMER
TAC60256 JTEXT ENV JMOR WAV TRS X FMASW 64db-24db 1994.11.25 17 A CUSTOMER
TAC60340 JTEXT ENV JMOR WAV TRS X FMEWA 56db-24db 1994.12.09 18 B CLERK
TAC60340 JTEXT ENV JMOR WAV TRS X FNOHI 64db-32db 1994.12.09 18 A CUSTOMER
TAC60382 JTEXT ENV JMOR WAV TRS X FKYIW 64db-24db 1994.12.15 19 A CUSTOMER
TAC70018 JTEXT ENV JMOR WAV TRS X FMISO 64db-32db 1994.00.00 28 A CUSTOMER
TAC70020 JTEXT ENV JMOR WAV TRS X FKEAO 64db-40db 1994.00.00 22 A CUSTOMER
TAC70021 JTEXT ENV JMOR WAV TRS X FCHNO 64db-24db 1994.00.00 17 B CLERK
TAC70024 JTEXT ENV JMOR WAV TRS X FMASH 64db-40db 1994.00.00 26 A CUSTOMER
TAC70025 JTEXT ENV JMOR WAV TRS X FSAAS 64db-40db 1994.00.00 11 A CUSTOMER
TAC70026 JTEXT ENV JMOR WAV TRS X FMAYO 64db-40db 1994.00.00 25 A CUSTOMER
TAC70104 JTEXT ENV JMOR WAV TRS X FAKTS 64db-32db 1994.00.00 22 A CUSTOMER
TAC71012 JTEXT ENV JMOR WAV TRS X FYOMO 64db-32db 1995.__.__ 15 A CUSTOMER
TAC71014 JTEXT ENV JMOR WAV TRS X FMAKS 66db-25db 1995.__.__ 17 A CUSTOMER
TAC71016 JTEXT ENV JMOR WAV TRS X FNAMU 59db-30db 1995.__.__ 24 A CUSTOMER
TAS70001 JTEXT ENV JMOR WAV TRS X FKAKA 64db-32db 1994.00.00 13 A CUSTOMER
TAS70002 JTEXT ENV JMOR WAV TRS X FSESE 64db-40db 1994.00.00 13 A CUSTOMER
TAS70003 JTEXT ENV JMOR WAV TRS X FSUSA 64db-16db 1994.00.00 17 B CLERK
TAS70005 JTEXT ENV JMOR WAV TRS X FTOCH 64db-48db 1994.00.00 12 A CUSTOMER
TAS70006 JTEXT ENV JMOR WAV TRS X FNOKU 72db-40db 1994.00.00 10 A CUSTOMER
TAS70007 JTEXT ENV JMOR WAV TRS X FYAMI 64db-40db 1994.00.00 13 A CUSTOMER
TAS70008 JTEXT ENV JMOR WAV TRS X FYUMI 72db-48db 1994.00.00 16 A CUSTOMER
TAS70012 JTEXT ENV JMOR WAV TRS X FAKMA 64db-40db 1994.00.00 14 A CUSTOMER
TAS70014 JTEXT ENV JMOR WAV TRS X FYUYA 64db-32db 1994.00.00 13 A CUSTOMER
TAC60270 JTEXT ENV JMOR WAV TRS X FTAKU 64db-24db 1994.11.30 10 A CUSTOMER
TAC60283 JTEXT ENV JMOR WAV TRS X FCHKY 56db-24db 1994.11.30 7 A CUSTOMER
TAC60325 JTEXT ENV JMOR WAV TRS X FNOKY 56db-32db 1994.12.09 9 A CUSTOMER
TAC60354 JTEXT ENV JMOR WAV TRS X FSHOO 64db-24db 1994.12.14 5 A CUSTOMER
TAC70302 JTEXT ENV JMOR WAV TRS X FCHAK 56db-32db 1994.00.00 19 A CUSTOMER
TAC71003 JTEXT ENV JMOR WAV TRS X FYUNV 57db-34db 1995.__.__ 12 A CUSTOMER
TAC71017 JTEXT ENV JMOR WAV TRS X FAKYY 61db-31db 1995.__.__ 8 A CUSTOMER
TAC71018 JTEXT ENV JMOR WAV TRS X FMITX 61db-27db 1995.__.__ 16 A CUSTOMER
```

付録 B Phoneme Set

The 26 phoneme label set that is given in Table 9 is used in all experiments (transcription and model files).

表 9: Phoneme set for Japanese speech recognition

label	example
a	akai (赤い)
b	benri (便利)
ch	chikara (力)
d	deguchi (出口)
e	eng (円)
g	gizhjutsu (技術)
h	hata (旗)
i	ima (今)
j	jama (山)
k	keqtei (決定)
m	mang (万)
n	namae (名)
ng	dengwa (電話)
o	otoko (男)
p	pang (パン)
q	aqsari (あっさり)
r	rongbung (論文)
s	sa (差)
sh	shjori (処理)
t	tani (谷)
ts	tsukau (使う)
u	ushi (牛)
w	dengwa (電話)
z	angzeng (安全)
zh	azhi (味)
-	pause or silence

付録 C Online Access to Recommended Acoustic Models

The recommended acoustic models, preprocessed parameter files, sample configuration files and some result files have been installed on `/dept1/work1/V1` (see Table 10). Additionally we received lexicon, language model and correct answer files for word recognition experiments from masataki@itl.atr.co.jp. We also installed parameter and answer files for test sets `TDMTa` and `TDMTb`, i.e. the conversation sets that are used for development of the integrated speech translation system. To facilitate comparisons, complete result files (lattice, n-best and additional information) for testsets `S1`, `S2` and `S4` have also been installed. The machines used were 500 MHz DEC Alpha (if you want to compare CPU time).

表 10: Directory tree for online access of model, parameter and result files

```
V1
 |- README
 |- model
 |   |- AM.MF.bin     # gender-independent A230/A230 CD800
 |   |- AM.M.bin      # male A230/M099 CD800
 |   |- AM.F.bin      # female A230/F131 CD800
 |   |- LEX.P         # lexicon for phoneme recognition
 |   |- LEX.W         # lexicon for word recognition
 |   |- LM.P          # language model for phoneme recognition
 |   |- LM.W          # language model for word recognition
 |   |- README
 |- config
 |   |- README
 |   |- config.P      # example config file for phoneme recognition
 |   |- config.W      # example config file for word recognition
 |   |- res.P
 |   |- res.W
 |- data
 |   |- README
 |   |- TAC70015.A.FSYNC    # framesync parameter file
 |   |- TAC70015.A.TRS      # correct answer for phoneme recognition
 |   |- TAC70015.A.ANS      # correct answer for word recognition
 |   |- TAC70015.A.sample   # input file for creating parameter file
...
 |   |- TSC71013.A.FSYNC
 |   |- TSC71013.A.TRS
 |   |- TSC71013.A.sample
 |- result
 |   |- AM.F.bin.r04a10.P   # female model, phoneme recognition
...
 |   |   |- TCS70023.A.res  # lattice and n-best results
 |   |   |- TCS70023.A.err  # timing information etc
 |   |   |- TCS70023.A.cnf  # actually used config file
...
 |   |- AM.F.bin.r04a10.W   # female model, word recognition
...
 |   |- AM.M.bin.r04a10.P
...
 |   |- AM.M.bin.r04a10.W
...
 |   |- AM.MF.bin.r04a10.P  # gender-independent model, phoneme recognition
...
 |   |- AM.MF.bin.r04a10.W
```

付録 D Detailed Recognition Results

More detailed results for the recommended models of the Jan97 experiments on a per speaker basis can be accessed for all speakers

Detailed results for all speakers can be found in Table 11. The summary information for male and female speakers are calculated considering all deletions, insertions and substitutions and not as averages of recognition accuracies.

Please note that in the word recognition experiments, 8 sentences were not scored for the gender independent acoustic models because the stacksize had been set too small (500,100). These sentences are shown in Table 12.

表 11: Detailed results by conversation for phoneme recognition and word recognition with the recommended acoustic models. Numbers given are accuracy of 1st best and network accuracy. Missing values exist either because of gender difference between speaker and model or because there existed no .JMOR file.

		phone			word		
conv	vers	M	F	MF	M	F	MF
TAC70015.A	r04a10	- -	66.0/75.7	64.2/78.1	- -	59.3/66.1	56.8/68.6
TAC70016.A	r04a10	69.0/78.4 - -		63.8/79.3	70.8/76.4 - -		66.0/73.6
TAC70017.A	r04a10	83.8/88.7 - -		74.9/86.2	81.0/83.5 - -		83.5/86.1
TAC70019.A	r04a10	- -	76.4/86.9	74.6/86.2	- -	76.7/79.7	80.5/85.0
TAC70021.A	r04a10	61.1/75.1 - -		55.5/70.4	55.3/64.2 - -		68.3/79.7
TAC70022.A	r04a10	67.6/78.2 - -		58.5/70.1	47.8/54.7 - -		56.1/65.5
TAC70023.A	r04a10	67.0/74.0 - -		64.5/76.3	62.4/68.8 - -		65.2/68.8
TAC70101.A	r04a10	- -	88.1/91.9	86.3/90.5	- -	86.5/87.9	88.7/90.8
TAC70102.A	r04a10	- -	78.3/88.2	75.3/86.9	- -	76.6/86.2	76.6/84.4
TAC70103.A	r04a10	71.5/81.6 - -		67.1/81.2	63.9/67.5 - -		72.3/78.3
TAC70201.A	r04a10	- -	79.6/88.4	79.4/89.0	- -	76.9/78.8	75.0/83.3
TAC70202.A	r04a10	74.7/84.6 - -		69.5/82.6	63.0/74.0 - -		55.2/65.2
TAC70203.A	r04a10	- -	74.7/83.3	69.0/80.4	- -	74.9/81.4	79.8/83.6
TAC70301.A	r04a10	- -	72.9/85.5	70.9/81.9	- -	77.6/85.1	75.4/82.1
TAC70303.A	r04a10	- -	87.5/91.5	85.6/90.4	- -	80.5/91.5	84.7/94.1
TAC70304.A	r04a10	47.1/62.8 - -		41.0/59.0	49.4/57.5 - -		71.2/83.3
TCC70103.A	r04a10	- -	77.6/86.1	74.0/83.3	- -	76.3/76.3	68.8/74.2
TCC70109.A	r04a10	73.7/83.2 - -		71.5/81.6	68.1/73.1 - -		62.2/73.9
TCC70201.A	r04a10	70.1/80.8 - -		66.3/77.8	46.6/52.4 - -		39.8/48.5
TCC70212.A	r04a10	64.8/75.6 - -		55.5/70.1	46.3/57.4 - -		40.5/52.5
TCC70307.A	r04a10	76.4/83.8 - -		74.1/84.9	73.9/78.2 - -		74.4/79.8
TCC71001.B	r04a10	- -	70.4/80.9	67.9/82.6	- - - -		- -
TCC71007.A	r04a10	- -	78.6/86.5	80.4/87.2	- - - -		- -
TCC71008.A	r04a10	73.0/82.2 - -		71.9/79.8	- - - -		- -
TCC71016.A	r04a10	- -	74.9/83.1	75.6/84.5	- - - -		- -
TCC71035.A	r04a10	46.6/66.2 - -		74.3/85.0	- - - -		- -
TCS70004.B	r04a10	- -	66.1/75.7	65.4/77.5	- -	57.6/65.3	61.0/72.9
TCS70010.A	r04a10	- -	72.9/79.8	69.8/78.0	- -	45.3/48.4	41.4/52.3
TCS70013.A	r04a10	- -	81.2/87.5	80.2/88.0	- -	74.6/84.2	74.6/81.6
TCS70020.A	r04a10	- -	41.3/58.4	63.3/76.2	- -	50.8/60.6	72.0/78.8
TCS70023.A	r04a10	- -	79.3/86.0	77.6/85.5	- -	76.3/83.7	70.0/76.7
TCS70025.A	r04a10	- -	79.3/85.9	76.7/86.3	- -	77.1/84.3	79.3/87.9
TCS70028.A	r04a10	- -	85.2/91.8	77.6/89.5	- -	84.5/87.3	85.5/89.1
TCS70034.A	r04a10	68.4/79.4 - -		66.9/80.2	44.0/56.0 - -		55.6/65.7
TCS70047.A	r04a10	- -	72.2/79.5	75.8/85.5	- -	67.4/77.0	74.1/83.7
TCS70055.A	r04a10	71.9/80.2 - -		71.2/82.4	59.5/67.9 - -		60.0/68.9
TCS70059.A	r04a10	- -	77.9/85.2	78.2/88.0	- -	78.0/85.0	77.0/79.0
TCS70070.A	r04a10	68.0/78.7 - -		65.7/76.3	52.9/62.4 - -		60.0/67.1
TCS70074.A	r04a10	81.9/87.8 - -		76.3/85.9	77.5/84.3 - -		82.4/85.3
TCS70082.A	r04a10	- -	74.4/83.9	71.7/83.7	- -	71.9/77.8	71.9/81.9
TSC71005.B	r04a10	- -	78.6/89.5	72.9/87.4	- - - -		- -
TSC71013.A	r04a10	- -	69.6/79.5	69.4/80.9	- - - -		- -
MALE	r04a10	69.0/79.2 - -		66.3/78.4	59.0/66.7 - -		61.6/69.9
FEMALE	r04a10	- -	74.5/83.3	73.5/84.0	- -	71.3/77.7	73.0/80.4

Example configuration files for phoneme recognition experiments and typewriter experiments are given in Tables 13 and 14. Important options are `lmscale` (language model scale), `beam` (beam width) and `work_area`. For details, please consult the ATRSPREC manual.

表 12: Ignored conversations due to too small stacksize for word recognition evaluation of gender independent acoustic model

```
AM.MF.bin.r04a10.W/TAC70022.A.res:comment=ERROR-UTTERANCE-10
AM.MF.bin.r04a10.W/TAC70304.A.res:comment=ERROR-UTTERANCE-1
AM.MF.bin.r04a10.W/TCC70212.A.res:comment=ERROR-UTTERANCE-2
AM.MF.bin.r04a10.W/TCC70307.A.res:comment=ERROR-UTTERANCE-2
AM.MF.bin.r04a10.W/TCS70004.B.res:comment=ERROR-UTTERANCE-7
AM.MF.bin.r04a10.W/TCS70004.B.res:comment=ERROR-UTTERANCE-10
AM.MF.bin.r04a10.W/TCS70004.B.res:comment=ERROR-UTTERANCE-12
AM.MF.bin.r04a10.W/TCS70034.A.res:comment=ERROR-UTTERANCE-16
```

表 13: Example configuration file of ATRlattice for phoneme recognition of conversation TAC70015.A

```
#I/Ocontrol config : Sun Mar  2 04:56:25 1997
        I/Ocontrol:rpcTable=
        I/Ocontrol:rpcNumber=3
        I/Ocontrol:outputByteorder=BigEndian
        I/Ocontrol:outputFd=stdout
        I/Ocontrol:outputParamType=
        I/Ocontrol:outputParamSize=
        I/Ocontrol:outputFormat=Lattice
        I/Ocontrol:inputByteorder=BigEndian
        I/Ocontrol:inputFd=/dept1/work1/V1/data/TAC70015.A.FSYNC
        I/Ocontrol:inputParamType=float
        I/Ocontrol:inputParamSize=34
        I/Ocontrol:inputFormat=FrameSync
#ATRresult config : Sun Mar  2 04:56:25 1997
        ATRresult:answer=/dept1/work1/V1/data/TAC70015.A.TRS
        ATRresult:dp_weight=1.0,1.0,1.0
        ATRresult:pause_symbol=-
        ATRresult:UTT_END=6
        ATRresult:UTT_START=5
        ATRresult:re_beam=
        ATRresult:N_best=1
        ATRresult:N_best_out=stdout
        ATRresult:lattice_out=stdout
#ATRlattice config : Sun Mar  2 04:56:25 JST 1997
        ATRlattice:lexicon=/dept1/work1/V1/model/LEX.P
        ATRlattice:amname=/dept1/work1/V1/model/AM.MF.bin
        ATRlattice:active_model=0
        ATRlattice:lmscale=4.000000,8.000000
        ATRlattice:wdpenalty=0,0
        ATRlattice:ngram=Class-2,/dept1/work1/V1/model/LM.P
        ATRlattice:beam=30,30
        ATRlattice:work_area=200,50
        ATRlattice:frame_shift=10
        ATRlattice:pause_symbol=-
        ATRlattice:dimension=34
        ATRlattice:phone_boundary=OFF
        ATRlattice:word_merge=non
        ATRlattice:UTT_START=5
        ATRlattice:UTT_END=6
        ATRlattice:backward_frame=-1
```

表 14: Example configuration file of ATRlattice for word recognition of conversation TAC70015.A

```
#I/Ocontrol config : Sun Mar  2 04:57:07 1997
        I/Ocontrol:rpcTable=
        I/Ocontrol:rpcNumber=3
        I/Ocontrol:outputByteorder=BigEndian
        I/Ocontrol:outputFd=stdout
        I/Ocontrol:outputParamType=
        I/Ocontrol:outputParamSize=
        I/Ocontrol:outputFormat=Lattice
        I/Ocontrol:inputByteorder=BigEndian
        I/Ocontrol:inputFd=/dept1/work1/V1/data/TAC70015.A.FSYNC
        I/Ocontrol:inputParamType=float
        I/Ocontrol:inputParamSize=34
        I/Ocontrol:inputFormat=FrameSync
#ATRresult config : Sun Mar  2 04:57:07 1997
        ATRresult:answer=/dept1/work1/V1/data/TAC70015.A.ANS
        ATRresult:dp_weight=1.0,1.0,1.0
        ATRresult:pause_symbol=-
        ATRresult:UTT_END=6
        ATRresult:UTT_START=5
        ATRresult:re_beam=
        ATRresult:N_best=1
        ATRresult:N_best_out=stdout
        ATRresult:lattice_out=stdout
#ATRlattice config : Sun Mar  2 04:57:07 JST 1997
        ATRlattice:lexicon=/dept1/work1/V1/model/LEX.W
        ATRlattice:amname=/dept1/work1/V1/model/AM.MF.bin
        ATRlattice:active_model=0
        ATRlattice:lmscale=8.000000,8.000000
        ATRlattice:wdpenalty=0,0
        ATRlattice:ngram=Class-2,/dept1/work1/V1/model/LM.W
        ATRlattice:beam=85,85
        ATRlattice:work_area=500,100
        ATRlattice:frame_shift=10
        ATRlattice:pause_symbol=-
        ATRlattice:dimension=34
        ATRlattice:phone_boundary=OFF
        ATRlattice:word_merge=non
        ATRlattice:UTT_START=5
        ATRlattice:UTT_END=6
        ATRlattice:backward_frame=-1
```

付録 E Implementation and Problems

All experiments have been implemented as a set of python scripts as enumerated in Table 15. Some of the problems, we found when implementing the experiments were:

- In general, this type of experiment is parallelizable on a coarse-grained level. However, dependencies are quite complicated. For example, once the topology has been calculated, the models for different number of states could be retrained in parallel on a cluster of workstations. For embedded reestimation we would also need the separately trained silence model. This type of dependency is difficult to express.
- Python can be written quite complicated. As I am adding more functions it becomes harder and harder to read.

表 15: Scripts used in the experiments

```
trainrec.py     # training and main
rec.py          # ATRlattice recognition
newscore.py     # scoring the results
spec.SDB.py     # defines most experimental parameters
```

付録 F Example HMnet Logfiles

The following two logfiles are examples for label training (see Table 16) and embedded reestimation (see Table 17) of the CD 800 A230/M099 models.

表 16: Logfile of label training for CD 800 A230/M099

```
;;; *************************************************************
;;; *             <<< SSS-ToolKit   Version 3.2 >>>             *
;;; *           Copyright(C) 1993                               *
;;; * ATR Interpreting Telecommunications Research Laboratories *
;;; *                       Department 1                        *
;;; *      2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02      *
;;; *                  Tel     : 07749-5-1301                   *
;;; *                  (Direct): 07749-5-1389                   *
;;; *                  Fax     : 07749-5-1308                   *
;;; *                  E-mail  : singer@itl.atr.co.jp             *
;;; *************************************************************
;; Thu Feb  6 21:12:21 1997
;;
;; <<< Retraining HMnet >>>
;;
;; Current Working Directory    : /tmp_mnt/dept1/work11/singer/RESULTS/33
;;
;; [HMnet]
;; Input Model File Name        : ModelMLSSS/Topology/HMnet_filled.5.800
;; State Number                 : 800
;; Tied Element                 : A230 M099
;; Factor Number                : 3
;; Main Element                 : a i k j o zh z u d m g ch ng r sh ts s e b q t w n p h
;; Remake Initial Parameter     : Yes
;; Training Type (FB/VIT/..)    : 0
;; Maximum BW Iteration         : 20
;; Output Model File Name       : ModelMLSSS/M099/HMnet_filled.5.800.tmp
;;
;; [Data]
;; Parameter Dimension          : 34
;; Total Data Number            : 37527
;; Total Sample Number          : 37527
;; Total Frame Length           : 300950
;; Min. Frame Length            : 4
;;
;; # Cut Data Number : 0
  0  1.153907e+07  0.000000e+00  5.745000e+01
  1  1.183659e+07  2.513600e-02  5.651667e+01
  2  1.193963e+07  8.630098e-03  5.625000e+01
  3  1.201094e+07  5.936809e-03  5.620000e+01
  4  1.206386e+07  4.386645e-03  5.620000e+01
  5  1.210327e+07  3.256287e-03  5.620000e+01
  6  1.213427e+07  2.554491e-03  5.615000e+01
  7  1.215794e+07  1.947628e-03  5.616667e+01
  8  1.217586e+07  1.471562e-03  5.620000e+01
  9  1.218987e+07  1.149515e-03  5.613333e+01
 10  1.220093e+07  9.060228e-04  5.611667e+01
 11  1.221014e+07  7.539742e-04  5.610000e+01
 12  1.221773e+07  6.215254e-04  5.611667e+01
 13  1.222459e+07  5.613123e-04  5.610000e+01
 14  1.223060e+07  4.913342e-04  5.608333e+01
 15  1.223583e+07  4.276070e-04  5.610000e+01
 16  1.224069e+07  3.967464e-04  5.608333e+01
 17  1.224468e+07  3.263496e-04  5.608333e+01
 18  1.224830e+07  2.955050e-04  5.608333e+01
 19  1.225183e+07  2.879240e-04  5.610000e+01
 20  1.225492e+07  2.519629e-04  2.136667e+01
# Iteration Times : 20
# Total Probability : 1.225492e+07
# CPU Time : 1264.383 sec
```

表 17: Logfile of embedded reestimation training for CD 800 A230/M099

```
;;; ************************************************************
;;; *             <<< SSS-ToolKit   Version 3.2 >>>            *
;;; *           Copyright(C) 1993                              *
;;; * ATR Interpreting Telecommunications Research Laboratories *
;;; *                        Department 1                      *
;;; *      2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02     *
;;; *                  Tel     : 07749-5-1301                  *
;;; *                  (Direct): 07749-5-1389                  *
;;; *                  Fax     : 07749-5-1308                  *
;;; *                  E-mail  : singer@itl.atr.co.jp             *
;;; ************************************************************
;; Thu Feb  6 21:39:23 1997
;;
;; <<< Retraining HMnet >>>
;;
;; Current Working Directory    : /tmp_mnt/dept1/work11/singer/RESULTS/33
;;
;; [HMnet]
;; Input Model File Name        : ModelMLSSS/M099/HMnet_filled.5.800
;; State Number                 : 801
;; Tied Element                 : M099 A230
;; Factor Number                : 3
;; Main Element                 : - a i k j o zh z u d m g ch ng r sh ts s e b q t w n p h
;; Remake Initial Parameter     : No
;; Training Type (FB/VIT/..)    : 1
;; Maximum BW Iteration         : 10
;; Output Model File Name       : ModelMLSSS/M099/HMnet_filled.emb.5.800
;;
;; [Data]
;; Parameter Dimension          : 34
;; Total Data Number            : 2348
;; Total Sample Number          : 53655
;; Total Frame Length           : 343135
;; Min. Frame Length            : 14
;;
;;  sample No.1344 is cut. ( path can not be found. )
;;  sample No.1724 is cut. ( path can not be found. )
;;  sample No.1745 is cut. ( path can not be found. )
;; # Cut Data Number : 3
  0  1.350513e+07  0.000000e+00  4.311167e+02
  1  1.380040e+07  2.139609e-02  4.317167e+02
  2  1.383335e+07  2.382031e-03  4.315167e+02
  3  1.385020e+07  1.216322e-03  4.313333e+02
  4  1.386088e+07  7.702383e-04  4.312333e+02
  5  1.386893e+07  5.809621e-04  4.311833e+02
  6  1.387531e+07  4.595872e-04  4.312333e+02
  7  1.388032e+07  3.606634e-04  4.312500e+02
  8  1.388452e+07  3.025928e-04  4.312167e+02
  9  1.388798e+07  2.491211e-04  4.311833e+02
 10  1.390213e+07  1.018090e-03  4.219167e+02
# Iteration Times : 10
# Total Probability : 1.390213e+07
# CPU Time : 4739.717 sec
```

付録 G Transcriptions Files (*.TRS)

An example for a transcription file (.TRS file) is given in Table 18. Start and end time for each chunk is given in msec. "-" is the silence label. (see also Appendix 付録 B).

表 18: Example for TRS file

```
560.00 sh,i,i,k,o,o,s,u,w,a 1430.00
1430.00 - 1970.00
1970.00 k,o,u,sh,i,n,o,s,u,t,e,e,k,i,n,i,n,a,q,t,e,o,r,i,m,a,s,u,g,a 3760.00
3760.00 - 4060.00
4060.00 o,o,d,o,b,u,r,u,w,a,i,k,a,g,a,n,a,s,a,i,m,a,s,u,k,a 5660.00
#
#
#
#
#
```

付録 H Experiments Dec96

Our initial set of experiments was essentially performed using the same conditions as the ones mentioned in Sections 2 to 4. However, there were not enough test speakers to get reliable results. For the sake of comparison we are adding these unreliable results.

H.1 Training and Test Sets Dec96

The testsets used were S1 and SL2, which are both dealing with room reservations. S1 (117 uttterances) is from the mono-lingual SDB database, SL2 (107 utterances) from the bilingual SLDB database. We thus used 9 female and 9 male speakers with a total of 224 utterances.

For training we used a total of 120 male and 160 female speakers mainly from the travel task and tried to exclude data from the scheduling task[8]. Some statistics for the training set are given in Table 19.

表 19: Details of training data sets (Dec96). Numbers in brackets are for phone segments longer or equal to 30 msec. Frameshift was 10 msec.

type	gender	#speakers	#pause units	#phones	#frames
M120	male	120	3011	66873 (53599)	427696 (395495)
F160	female	160	4192	97081 (78834)	653242 (604323)
A280	both	280	7203	163954 (132433)	1080938 (999818)

The 34-dimensional parameter files took 83,403,148 byte for F160 and 138,015,797 byte for A280.

H.2 Experimental Results Dec96

Phoneme accuracy results for testsets S1 and SL2 using differently trained models are given in Tables 20 and 21.

表 20: Phoneme accuracy for male test speakers (Dec96): "CI X" stands for context independent, X mixtures left-to-right, 3 states HMM's. "CD X" stands for context dependent, X states, 5 mixture HMnet.

type	#gaussians	A280/A280	M120/M120	A280/M120	F160/M120
CI 5	385	43.4	48.8	NA	NA
CI 10	760	47.6	50.8	NA	NA
CI 15	1135	45.8	53.4	NA	NA
CD 400	2010	58.7	64.4	-	-
CD 600	3010	62.6	67.1	66.1	64.7
CD 800	4010	64.4	66.9	-	-

表 21: Phoneme accuracy for female test speakers (Dec96): "CI X" stands for context independent, X mixtures left-to-right, 3 states HMM's. "CD X" stands for context dependent, X states, 5 mixture HMnet. Numbers in brackets are for a "failed" experiment, where only about 120 female conversations were used.

type	#gaussians	A280/A280	F160/F160	A280/F160	M120/F160
CI 5	385	46.6	53.4 (53.7)	NA	NA
CI 10	760	50.4	55.3 (56.3)	NA	NA
CI 15	1135	50.5	55.1 (56.3)	NA	NA
CD 400	2010	68.4	71.0 (71.7)	-	-
CD 600	3010	73.4	73.9 (75.6)	75.3	73.4
CD 800	4010	74.9	74.5 (76.3)	-	-

A280/M120 stands for topology training with dataset A280 and parameter reestimation (label and embedded) with dataset M120. We only calculated all combinations for the CD600 case.

(H.2.1) Discussion

For both, male and female testsets, the accuracy increases as the number of states is increased. The only exception is CD800 for male speakers.

One of the most interesting results was F160/F160, where initially we had a a "failed" experiment, where only about 120 female conversations were used, both for topology training and reestimation. The results were about 1.5 % better than with the full 160 speakers. This suggests that the choice of trainingset is important. We also should check the transcriptions more thoroughly.

Additionally we ran experiments on SL3, i.e. a closed language model testset. We compared different segmentation schemes and versions of the software (thanks to shimizu@itl.atr.co.jp). Acoustic models used were F160/F160 and M120/M120, i.e. "pure" gender dependent models (see Table 22).

The results indicate that the newly trained models are performing about 5% better than the previously used "standard" acoustic models. The choice of segmentation can have a huge influence on recognition performance, e.g. for TAS22001.

表 22: Comparison for two "extreme" conversations in testset SL3. number in brackets are network accuracy

	r04a07		r04a04	r04a07
	segmentation(c)	segmentation(b)		
	new model			old model
TAS12001	65.7 (77.7)	68.1 (80.1)	68.9 (69.4)	64.2 (77.9)
TAS22001	49.6 (67.1)	69.1 (87.1)	70.0 (72.9)	65.7 (83.2)