

Incorporating Knowledge Sources into Statistical Speech Recognition

Lecture Notes in Electrical Engineering

Incorporating Knowledge Sources into Statistical Speech Recognition
Sakti, Sakriani, Markov, Konstantin, Nakamura, Satoshi, and Minker, Wolfgang
978-0-387-85829-6

Intelligent Technical Systems
Martínez Madrid, Natividad; Seepold, Ralf E.D. (Eds.)
978-1-4020-9822-2

Languages for Embedded Systems and their Applications
Radetzki, Martin (Ed.)
978-1-4020-9713-3

Multisensor Fusion and Integration for Intelligent Systems
Lee, Sukhan; Ko, Hanseok; Hahn, Hernsoo (Eds.)
978-3-540-89858-0

Designing Reliable and Efficient Networks on Chips
Murali, Srinivasan
978-1-4020-9756-0

Trends in Communication Technologies and Engineering Science
Ao, Sio-long; Huang, Xu; Wai, Ping-kong Alexander (Eds.)
978-1-4020-9492-7

Functional Design Errors in Digital Circuits: Diagnosis Correction and Repair
Chang, Kai-hui, Markov, Igor, Bertacco, Valeria
978-1-4020-9364-7

Traffic and QoS Management in Wireless Multimedia Networks: COST 290 Final Report
Koucheryavy, Y., Giambene, G., Staehle, D., Barcelo-Arroyo, F., Braun, T., Siris, V. (Eds.)
978-0-387-85572-1

Proceedings of the 3rd European Conference on Computer Network Defense
Siris, V.; Ioannidis, S.; Anagnostakis, K.; Trimintzios, P. (Eds.)
978-0-387-85554-7

Intelligentized Methodology for Arc Welding Dynamical Processes: Visual Information Acquiring,
Knowledge Modeling and Intelligent Control
Chen, Shan-Ben, Wu, Jing
978-3-540-85641-2

Proceedings of the European Computing Conference: Volume 2
Mastorakis, Nikos, Mladenov, Valeri, Kontargyri, Vassiliki T. (Eds.)
978-0-387-84818-1

Proceedings of the European Computing Conference: Volume 1
Mastorakis, Nikos, Mladenov, Valeri, Kontargyri, Vassiliki T. (Eds.)
978-0-387-84813-6

Electronics System Design Techniques for Safety Critical Applications
Sterpone, Luca
978-1-4020-8978-7

Data Mining and Applications in Genomics
Ao, Sio-long
978-1-4020-8974-9

Continued after index

Sakriani Sakti • Konstantin Markov •
Satoshi Nakamura • Wolfgang Minker

Incorporating Knowledge Sources into Statistical Speech Recognition

Sakriani Sakti
NICT/ATR Spoken Language
Communication Research Laboratories
Keihanna Science City
Kyoto, Japan

Konstantin Markov
NICT/ATR Spoken Language
Communication Research Laboratories
Keihanna Science City
Kyoto, Japan

Satoshi Nakamura
NICT / ATR Spoken Language
Communication Research Laboratories
Keihanna Science City
Kyoto, Japan

Wolfgang Minker
University of Ulm
Ulm, Germany

ISBN 978-0-387-85829-6 e-ISBN 978-0-387-85830-2
DOI: 10.1007/978-0-387-85830-2

Library of Congress Control Number: 2008942803

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of going to press, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper.

springer.com

*This book is dedicated
to our parents and families
for their support and endless love*

Preface

State-of-the-art automatic speech recognition (ASR) systems use statistical data-driven methods based on hidden Markov models (HMMs). Although such approaches have proved to be efficient choices, ASR systems often perform much worse than human listeners, especially in the presence of unexpected acoustic variability. To improve performance, we usually rely on collecting more data to train more detailed models. However, such resources are rarely available, since the presence of variabilities in speech arise from many different factors, and thus a huge amount of training data is required to cover all possible variabilities. In other words, it is not enough to handle these variabilities by relying solely on statistical models. The systems need additional knowledge on speech that could help to handle these sources of variability. Otherwise, only a limited level of success could be achieved.

Many researchers are aware of this problem, and thus various attempts to integrate more explicitly knowledge-based and statistical approaches have been made. However, incorporating various additional knowledge sources often leads to a complicated model, where achieving optimal performance is not feasible due to insufficient resources or data sparseness. As a result, input space resolution may be lost due to non-robust estimates and the increased number of unseen patterns. Moreover, decoding with large models may also become cumbersome and sometimes even impossible.

This book addresses the problem of developing efficient ASR systems that can maintain a balance between utilizing wide-ranging knowledge of speech variability while keeping the training/recognition effort feasible, of course while also improving speech recognition performance. In this book, an efficient general framework to incorporate additional knowledge sources into state-of-the-art statistical ASR systems is provided. It can be applied to many existing ASR problems with their respective model-based likelihood functions in flexible ways.

Since there are various types of knowledge sources from different domains, it may be difficult to formulate a probabilistic model without learning the dependencies between the sources. To solve such problems in a unified way, the

work reported in this book adopts the Bayesian network (BN) framework. This approach allows the probabilistic relationship between information sources to be learned. Another advantage of the BN framework lies in the fact that it facilitates the decomposition of the joint probability density function (PDF) into a linked set of local conditional PDFs based on the junction tree algorithm. Consequently, a simplified form of the model can be constructed and reliably estimated using a limited amount of training data.

This book focuses on the acoustic modeling problem as arguably the central part of any speech recognition system. The incorporation of various knowledge sources, including background noises, accent, gender and wide phonetic knowledge information, in modeling is also discussed. Such an application often suffers from a sparseness of data and memory constraints. First, the additional sources of knowledge are incorporated at the HMM state distribution. Then, these additional sources of knowledge are incorporated at the HMM phonetic modeling. The presented approaches are experimentally verified in the large-vocabulary continuous-speech recognition (LVCSR) task. The book closes with a summary of the described methods and the results of the evaluations.

Contents

1	Introduction and Book Overview	1
1.1	Automatic Speech Recognition - A Way of Human-Machine Communication	1
1.2	Approaches to Speech Recognition	4
1.2.1	Knowledge-based Approaches	4
1.2.2	Corpus-based Approaches	6
1.3	State-of-the-art ASR Performance	7
1.4	Studies on Incorporating Knowledge Sources	10
1.4.1	Sources of Variability in Speech	10
1.4.2	Existing Ways of Incorporating Knowledge Sources	12
1.4.3	Major Challenges to Overcome	15
1.5	Book Outline	16
2	Statistical Speech Recognition	19
2.1	Pattern Recognition Overview	19
2.2	Theory of Hidden Markov Models	22
2.2.1	Markov Chain	22
2.2.2	General form of an HMM	23
2.2.3	Principle Cases of HMM	25
2.3	Pattern Recognition for HMM-Based ASR Systems	35
2.3.1	Front-end Feature Extraction	36
2.3.2	HMM-Based Acoustic Model	43
2.3.3	Pronunciation Lexicon	49
2.3.4	Language Model	50
2.3.5	Search Algorithm	51
3	Graphical Framework to Incorporate Knowledge Sources	55
3.1	Graphical Model Representation	56
3.1.1	Probability Theory	56
3.1.2	Graphical Model	59
3.1.3	Junction Tree Algorithm	63

3.2	Procedure of GFIKS	68
3.2.1	Causal Relationship between Information Sources	70
3.2.2	Direct Inference on Bayesian Network	71
3.2.3	Junction Tree Decomposition	72
3.2.4	Junction Tree Inference	75
3.3	Practical Issues of GFIKS	75
3.3.1	Types of Knowledge Sources	75
3.3.2	Different Levels of Incorporation	76
4	Speech Recognition Using GFIKS	79
4.1	Applying GFIKS at the HMM State Level	79
4.1.1	Causal Relationship Between Information Sources	80
4.1.2	Inference	81
4.1.3	Enhancing Model Reliability	81
4.1.4	Training and Recognition Issues	82
4.2	Applying GFIKS at the HMM Phonetic-unit Level.....	83
4.2.1	Causal Relationship between Information Sources	83
4.2.2	Inference	85
4.2.3	Enhancing the Model Reliability	85
4.2.4	Deleted Interpolation	86
4.2.5	Training and Recognition Issues	86
4.3	Experiments with Various Knowledge Sources	87
4.3.1	Incorporating Knowledge at the HMM State Level.....	87
4.3.2	Incorporating Knowledge at the HMM Phonetic-unit Level	116
4.4	Experiments Summary and Discussion	132
5	Conclusions and Future Directions	139
5.1	Conclusions	139
5.1.1	Theoretical Issues	139
5.1.2	Application Issues	140
5.1.3	Experimental Issues	141
5.2	Future Directions: A Roadmap to a Spoken Language Dialog System	142
A	Speech Materials	145
A.1	AURORA TIDigit Corpus	145
A.2	TIMIT Acoustic-Phonetic Speech Corpus	146
A.3	Wall Street Journal Corpus	148
A.4	ATR Basic Travel Expression Corpus	150
A.5	ATR English Database Corpus	150

B	ATR Software Tools	153
B.1	Generic Properties of ATRASR	153
B.2	Data Preparation	153
B.3	SSS Data Generating Tools	155
B.4	Acoustic Model Training Tools	155
B.5	Language Model Training Tools	157
B.6	Recognition Tools	157
C	Composition of Bayesian Wide-phonetic Context	163
C.1	Proof using Bayes's Rule	163
C.2	Variants of Bayesian Wide-phonetic Context Model	164
D	Statistical Significance Testing	169
D.1	Statistical Hypothesis Testing	169
D.2	The Use of the Sign Test for ASR	172
	References	175
	Index	189

List of Figures

1.1	A machine that recognizes the speech waveform of a human utterances as “Good night.”	2
1.2	Knowledge-based ASR system.	4
1.3	Speech spectrogram reading, which corresponds to the word sequence of “Good night”	5
1.4	Corpus-based statistical ASR system.	6
1.5	2003 NIST’s benchmark ASR test history (After Pallett, 2003, ©2003 IEEE).	7
1.6	TC-STAR ASR evaluation campaign (After Choukri, 2007, ©TC-STAR).	8
1.7	“S” curve of ASR technology progress and the predicted performance from combining deep knowledge with a statistical approach.	9
1.8	Incorporating knowledge into a corpus-based statistical ASR system.	16
2.1	Pattern recognition: Establishing mapping from multi-dimensional measurement space X to three-class target decision space Y .	20
2.2	Pattern recognition approach for ASR: Establish mapping from measurement space X of speech signal to target decision space Y of word strings.	22
2.3	Simple three-state Markov chain for daily weather.	22
2.4	HMM of the daily weather, where there is no deterministic meaning on any state.	24
2.5	Left-to-right HMM of the daily weather.	25
2.6	Process flow on trellis diagram of 3-state HMM with time length T .	26
2.7	Forward probability function representation (for $j=1$).	27
2.8	Backward probability function representation (for $i=1$).	28

2.9	Example of finding the best path on a trellis diagram using the Viterbi algorithm.	30
2.10	Graphical interpretation of the EM algorithm.	31
2.11	Forward-backward probability function representation.	32
2.12	A generic automatic speech recognition system, composed of five components: feature extraction, acoustic model, pronunciation lexicon, language model and search algorithm.	36
2.13	Source-Filter model of the speech signal $x[n] = e[n] * h[n]$	37
2.14	Source-filter separation by cepstral analysis.	37
2.15	(a) A windowed speech waveform. (b) The spectrum of Figure 2.15(a). (c) The resulting cepstrum. (d) The Fourier transform of the low-frequency component.	38
2.16	MFCC feature extraction technique, which generates a 25-dimensional feature vector x_t for each frame.	41
2.17	A summary of feature extraction process, producing a feature vector which correspond to one point in a multi-dimensional space.	43
2.18	Discrete HMM observation density where the emission statistics or HMM state output probabilities are represented by discrete symbols.	44
2.19	Continuous GMM, where the continuous observation space is modeled using mixture Gaussians (state-specific). They are weighted and added to compute the emission statistic likelihoods (HMM state output probabilities).	45
2.20	Structure example of the monophone $/a/$ HMM acoustic model.	46
2.21	Structure example of the triphone $/a^-, a, a^+ /$ HMM acoustic model.	47
2.22	Shared-state structures of the triphone $/a^-, a, a^+ /$ HMM acoustic model.	47
2.23	An example of a phonetic decision tree for HMM state of the triphone with the central phoneme $/ay/$	48
2.24	Contextual splitting and temporal splitting of SSS algorithm (After Jitsuhiro, 2005).	49
2.25	Example of a tree-based pronunciation lexicon.	50
2.26	Multi-level probability estimation of statistical ASR.	52
3.1	Incorporating knowledge into corpus-based statistical ASR system.	55
3.2	Two equivalent models that can be obtained from each other through arc reversal of Bayes's rule, since $P(a,b)=P(b,a)$	60
3.3	Graphical representation of $P(a b_1, b_2, ..., b_n)$	60
3.4	Three BNs with different arrow directions over the same random variables a , b , and c . They appear in the case of serial, diverging, and converging connection, respectively.	61

3.5	Example of BN topology describing conditional relationship among a, b, c, d, e, f, g and h	63
3.6	Moral and triangulated graph of Figure 3.5.	64
3.7	Junction graph of Figure 3.5.	66
3.8	The resulting junction tree.	66
3.9	Clique $C_1 = [a, b, d]$ in the original graph of Figure 3.5.	67
3.10	General procedure of GFIKS (graphical framework to incorporate additional knowledge sources).	69
3.11	(a) BN topology describing the conditional relationship between data D and model M . (b) BN topology describing the conditional relationship among D, M , and additional knowledge K	70
3.12	Examples of BN topologies describing the conditional relationship among data D , model M , and several knowledge sources K_1, K_2, \dots, K_N	71
3.13	(a) BN topology describing conditional relationship among D, M, K_1 , and K_2 . (b) Moral and triangulated graph of Figure 3.13(a). (c) Equivalent BN topology. (d) Moral and triangulated graph of Figure 3.13(c). (e) Junction tree of Figure 3.13(d).	73
3.14	(a) Equivalent BN topology of the BN shown in Figure 3.12(a). (b) Corresponding junction tree.	74
3.15	Incorporating knowledge sources into HMM state (denoted by a small box) and phonetic unit level (denoted by a large box).	77
4.1	(a) Applying GFIKS at the HMM state level. (b) BN topology structure describing the conditional relationship between HMM state Q and observation vector X	80
4.2	BN topology structure after incorporating additional knowledge sources K_1, K_2, \dots, K_N in HMM state distribution $P(X, Q)$ (assuming that all K_1, K_2, \dots, K_N are independent given Q).	81
4.3	Example of observation space modeling by BN, where each value of K_i corresponds to a different Gaussian.	82
4.4	(a) Applying GFIKS at the HMM phonetic-unit level. (b) BN topology structure describing the conditional relationship between HMM phonetic model λ and observation segment X_s	84
4.5	BN topology structure after incorporating additional knowledge sources K_1, K_2, \dots, K_N in HMM phonetic model $P(X_s, \lambda)$ (assuming that all K_1, K_2, \dots, K_N are independent given λ).	84
4.6	Rescoring procedure with the composition models.	87
4.7	BN topology structure showing the conditional relationship among HMM state Q , observation vector X , and additional knowledge source of gender information G	88

4.8	Recognition accuracy rates of proposed HMM/BN, which are comparable with those of other systems from the “Hub and Spoke Paradigm for Continuous Speech Recognition Evaluation” for primary condition of WSJ Hub2-5k task.	93
4.9	BN topology structure describing the conditional relationship between HMM state Q , observation vector X , and additional knowledge sources of noise type N and SNR value S	94
4.10	Comparison of different systems: HMM, DBN (Bilmes et al., 2001), and proposed HMM/BN	98
4.11	BN topologies of the left state (a), center state (b), and right state (c) of LR-HMM/BN for modeling a pentaphone context $/a^{--}, a^-, a, a^+, a^{++}/$	99
4.12	BN topologies of the left state (a), center state (b), and right state (c) of LRC-HMM/BN, for modeling a pentaphone context $/a^{--}, a^-, a, a^+, a^{++}/$	100
4.13	Observation space modeling by BN, where a different value of second following context C_R corresponds to a different Gaussian.	101
4.14	Knowledge-based phoneme classes of the observation space.	102
4.15	Determining distance metric by Euclidean distance.....	103
4.16	Data-driven phoneme classes of observation space.	103
4.17	Recognition accuracy rates of pentaphone LR-HMM/BN using knowledge-based second preceding and following context clustering.	106
4.18	Recognition accuracy rates of pentaphone LRC-HMM/BN using knowledge-based second preceding and following context clustering.	107
4.19	Recognition accuracy rates of pentaphone LR-HMM/BN and LRC-HMM/BN using data-driven Gaussian clustering.	108
4.20	Comparing recognition accuracy rates of triphone HMM and pentaphone HMM/BN models with a fixed and a varied number of mixture components per state, but having the same 15 mixture components per state on average.	109
4.21	Topology of fLRC-HMM/BN for modeling a pentaphone context $/a^{--}, a^-, a, a^+, a^{++}/$, where state PDF has additional variables C_L and C_R representing the second preceding and following contexts, respectively.	110
4.22	(a) fLRCG-HMM/BN topology with additional knowledge G , C_L and C_R , (b) fLRCA-HMM/BN topology with additional variables A , C_L , and C_R , and (c) fLRCAG-HMM/BN topology with additional knowledge A , G , C_L , and C_R	111
4.23	Recognition accuracy rates of proposed HMM/BN models having identical numbers of 5, 10, and 20 mixture components per state.....	113

4.24	Comparing recognition accuracy rates of different systems: triphone HMM baseline, pentaphone HMM baseline, and the proposed pentaphone HMM/BN models having the same five mixture components per state.	115
4.25	BN topology structure describing the conditional relationship among X_s , λ , C_L , and C_R	116
4.26	(a) Equivalent BN topology. (b) Moral and triangulated graph of Figure 4.26(a). (c) Junction tree of Figure 4.26(b).	117
4.27	(a) Conventional triphone model, (b) Conventional pentaphone model, (c) Bayesian pentaphone model composition C1L3R3, consisting of the preceding/following triphone-context unit and center-monophone unit.	119
4.28	Rescoring procedure with pentaphone composition models: C1L3R3 or C3L4R4.	120
4.29	N-best rescoring mechanism.	121
4.30	Recognition accuracy rates of Bayesian triphone model.	122
4.31	Recognition accuracy rates of Bayesian pentaphone models.	124
4.32	Relative reductions in WER by Bayesian triphone C1L2R2 model from monophone baseline and by Bayesian pentaphone C1L3R3 model from triphone baseline.	125
4.33	Recognition accuracy rates of conventional pentaphone C5 and proposed Bayesian pentaphone C1L3R3 models with different amounts of training data.	126
4.34	BN topology structure describing the conditional relationship among X_s , λ , C_L , C_R , A , and G	127
4.35	(a) Equivalent BN topology of Figure 4.34. (b) Moral and triangulated graph of Figure 4.35(a). (c) Corresponding junction tree.	127
4.36	Rescoring procedure with the accent-gender-dependent pentaphone composition models: C1L3R3, C1L3R3-A, C1L3R3-G, and C1L3R3-AG.	128
4.37	Comparing recognition accuracy rates of different systems triphone HMM baseline, pentaphone HMM baseline, and proposed pentaphone models having the same 5, 10, and 20 mixture components per state.	130
4.38	Comparing recognition accuracy rates of different systems: triphone HMM baseline, pentaphone HMM baseline, and proposed models incorporating knowledge sources at HMM state and phonetic unit levels.	136
5.1	Roadmap to spoken language dialog system incorporating other knowledge sources at higher ASR levels.	143
B.1	The ATRASR phoneme-based SSS data creation for phone-unit model training.	156

- B.2 The ATRASR topology training for each phone acoustic-unit model. 158
- B.3 The ATRASR embedded training for a whole HMnet. 159
- B.4 The recognition process using ATRASR tools. 160

- C.1 Bayesian pentaphone model composition. (a) is C5, the conventional pentaphone model, (b) is Bayesian C1L3R3, which is composed of the preceding/following triphone-context unit and center-monophone unit, (c) is Bayesian C3L4R4, which is composed of the preceding/following tetraphone-context unit and center-triphone-context unit, (d) is Bayesian C1Lsk3Rsk3, which is composed of the preceding/following skip-triphone-context unit and center-monophone unit, and (e) is Bayesian C1C3Csk3, which is composed of the center skip-triphone-context unit, center triphone-context unit and center-monophone unit. 167

- D.1 The distribution of population according to the null hypothesis (H_0 is true), with upper-tail of rejection region for $P \leq \alpha$ 171

List of Tables

4.1	<i>English phoneme set.</i>	90
4.2	<i>1993 Hub and Spoke CSR evaluation on Hub 2: 5k read WSJ task (Kubala et al., 1994; Pallett et al., 1994).</i>	92
4.3	<i>HMM/BN system performance on Hub 2: 5k read WSJ task.</i>	93
4.4	<i>Recognition accuracy rates (%) for proposed HMM/BN on AURORA2 task.</i>	97
4.5	<i>Knowledge-based phoneme classes based on manner of articulation.</i>	101
4.6	<i>Recognition accuracy rates (%) for proposed pentaphone HMM/BN model using fLRC-HMM/BN (see Figure 4.22) on a test set of matching accents with different numbers of mixture components.</i>	114
4.7	<i>Recognition accuracy rates (%) for proposed pentaphone HMM/BN model using fLRC-HMM/BN (see Figure 4.22) on a test set of mismatched accents with 15 mixture components.</i>	115
4.8	<i>Recognition accuracy rates (%) for proposed Bayesian pentaphone C1L3R3-AG (see Eq. (4.30)) on a test set of matching accents with different numbers of mixture components.</i>	131
4.9	<i>Recognition accuracy rates (%) for proposed Bayesian pentaphone C1L3R3-AG model (see Eq. (4.30)) on a test set of mismatched accents with 15 mixture components.</i>	132
4.10	<i>Summary of incorporating various knowledge sources at the HMM state level.</i>	134
4.11	<i>Summary of incorporating various knowledge sources at the HMM phonetic unit level.</i>	135
A.1	<i>Dialect distribution of speakers.</i>	147
A.2	<i>Speech materials of TIMIT database.</i>	148
A.3	<i>Statistics on the TIMIT database.</i>	148
A.4	<i>Text sentence materials of ATR English speech database.</i>	151
A.5	<i>Speech materials of ATR English speech database.</i>	151

Glossary

AM	Acoustic model
ARPA	Advanced Research Projects Agency
ASR	Automatic speech recognition
A-STAR	Asian speech translation advanced research
ATR	Advanced Telecommunication Research
AUS	Australian
BN	Bayesian network
BRT	British
BTEC	Basic travel expression corpus
BU	Boston University
C1	Center monophone unit
C3	Center triphone context
Csk3	Center skip-triphone context
C5	Center pentaphone context
CCCC	CSR corpus coordinating committee
CNRS-LIMSI	France's National Center for Scientific Research
CPD	Conditional probability distribution
CPT	Conditional probability table
CSR	Continuous speech recognition
C-STAR	Consortium for speech translation advanced research
CU	Cambridge University
DAG	Directed acyclic graph
DARPA	Defense Advanced Research Projects Agency
DBN	Dynamic Bayesian network
DCT	Discrete cosine transform
DEL	Deletions
DI	Deleted interpolation
DSR	Distributed speech recognition
EDB	English database
ELRA	European language resources association

XXII GLOSSARY

EM	Expectation-maximization
EPPS	European Parliament Plenary Sessions
fLRC-HMM/BN	Full HMM/BN for left, right and center state
fLRCA-HMM/BN	Full HMM/BN for left, right and center state, including accent dependency
fLRCAG-HMM/BN	Full HMM/BN for left, right and center state, including accent and gender dependency
fLRG-HMM/BN	Full HMM/BN for left, right and center state, including gender dependency
FFT	Fast Fourier transform
GDHMM	Gender-dependent Hidden Markov model
GFIKS	Graphical framework to incorporate additional knowledge sources
GIHMM	Gender-independent Hidden Markov model
GMM	Gaussian mixture model
HMM	Hidden Markov model
ICASSP	International conference on acoustics, speech and signal processing
ICSI	International Computer Science Institute
ICSLP	International conference on spoken language processing
IEEE	Institute of Electrical and Electronics Engineers
IEICE	Institute of Electronics, Information and Communication Engineers
Imp	Improvement
INS	Insertions
L3	Left triphone context
L4	Left tetraphone context
LM	Language model
LPC	Linear prediction coefficients
LRC-HMM/BN	HMM/BN for left, right and center state
LR-HMM/BN	HMM/BN for left and right state
Lsk3	Left skip-triphone context
LVCSR	Large-vocabulary continuous-speech recognition
MAD	Machine translation aided dialogue
MAP	Maximum <i>a posteriori</i>
MDL	Minimum description length
MFCC	Mel-frequency cepstral coefficients
MIT	Massachusetts Institute of Technology
ML	Maximum likelihood
MLLR	Maximum likelihood linear regression
MSG	Modulation-filtered spectrogram
MT	Machine translation
NIST	National Institute of Standards and Technology
NOVO	Noise voice composition
PDF	Probability density function

PLP	Perceptual linear prediction
PMC	Parallel model combination
R3	Right triphone context
R4	Right tetraphone context
Rel	Relative
Resc	Rescoring
Rsk3	Right skip-triphone context
S2ST	Speech-to-speech translation
SD	Speaker dependent
SI	Speaker independent
SIL	Silence
SLC	Spoken Language Communication
SNR	Signal-to-noise ratio
SSS	Successive state splitting
STQ	Speech processing, transmission and quality
SUB	Substitutions
SWB	Switchboard
TC-STAR	Technology and corpora for speech to speech translation research
TI	Texas Instrument
US	United States
VQ	Vector quantization
WER	Word error rate
WFST	Weighted finite state transducers
WSJ	Wall Street journal