A METHOD TO INTEGRATE ADDITIONAL KNOWLEDGE SOURCES INTO HMM BASED ON JUNCTION TREE DECOMPOSITION

Sakriani Sakti[1,2], *Konstantin Markov*[1,2], *Satoshi Nakamura*[1,2]

[1]National Institute of Information and Communications Technology, Japan
[2]ATR Spoken Language Communication Research Laboratories, Japan
{sakriani.sakti,konstantin.markov,satoshi.nakamura}@atr.jp

ABSTRACT

Most current 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 still perform much worse than human listeners, especially in the presence of unexpected acoustic variability. Only a limited level of success can be achieved, by relying only on statistical models and mostly ignoring the additional knowledge available. We propose a new method of integrating various kinds of additional knowledge sources into an HMM-based statistical acoustic model in this paper. We utilized the junction tree algorithm to achieve efficient integration due to increased model complexity. This is since it facilitates the decomposition of the joint probability density function (PDF) into a linked set of local conditional PDFs. This way, a simplified form of the model could be constructed and reliably estimated using limited training data. We evaluated how efficient the proposed method was on an LVCSR task using two different types of accented English speech data. The experimental results revealed that our method improved word accuracy with respect to the standard HMM.

1. INTRODUCTION

Numerous researchers have worked in the area of ASR for about the past four decades. The promise is to develop an intelligent machine that can automatically recognize naturally spoken words uttered by humans. However, extracting the underlying linguistic message from a complex acoustic signal is not an easy task, due to many sources of variability that are contained in the signal [1].

Several approaches have been developed to address this problem. The approaches to ASR can generally be classified into two main types: ”knowledge-based” and ”corpus-based”. The former is mainly based on human ability to interpret spectrograms or other visual representations of the speech signal by knowledge-based rules [2, 3, 4]. However, there are underlying problems in the fact that it is difficult to envisage all possible ways in which these rules are interdependent. Consequently, some rules inevitably compete with others that explain the same phenomenon while still others are in direct contradiction [5]. The latter approach, in contrast, is usually based on modeling the speech signal using well-defined statistical algorithms that can automatically extract knowledge from the data. This approach has achieved encouraging results, and outperforms the previous knowledge-based approach. That is why most current ASR systems usually use statistical data-driven methods based on hidden Markov models (HMMs). Today’s state-of-the-art ASR systems attain very good performance in controlled conditions. Although such approaches have proved to be efficient choices, ASR systems still often perform much worse than human listeners, especially in the presence of unexpected acoustic variability. Only a limited level of success can be achieved, by relying only on statistical models and mostly ignoring the additional knowledge available.

Various attempts to integrate more explicit knowledge-based and statistical approaches have also been undertaken. For example, [6] proposed that acoustic phonetic knowledge sources be incorporated using neural networks. Other works such as [7, 8] proposed that articulatory features, sub-band correlation, or speaking styles be incorporated by utilizing dynamic Bayesian networks (DBNs). However, there are often cases when developing such complex models and achieving optimal performance is not feasible. These occur especially when the resources, i.e. available training data and memory space, are inadequate to properly train the model parameters. Input space resolution may be lost as a result due to non-robust estimates and an increased number of unseen patterns. Moreover, decoding using a large model may also become cumbersome and sometimes even impossible. The best we can do is to choose a simplified form of the model that can be reliably estimated using the training data available.

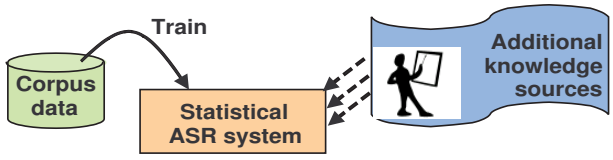

Figure 1: Integrating knowledge sources into statistical ASR system

We propose a new method of integrating various kinds of additional knowledge sources into an HMM-based statistical acoustic model in this paper, as outlined in Fig 1. We utilized the junction tree algorithm to achieve efficient integration due to increased model complexity. This is since it facilitates the decomposition of the joint probability density function (PDF) into a linked set of local conditional PDFs. This way, a simplified form of the model can be constructed and reliably estimated using limited training data.

We will first explain how we applied the proposed framework to the problem of integrating wide-phonetic knowledge information, which often suffers from data sparsity and memory constraints. We then attempted to integrate other additional knowledge, such as accent and gender information. The performance of the proposed model was experi-

mentally demonstrated in an LVCSR task using accented English speech data.

The next section describes the general framework for incorporating additional sources of knowledge and details about junction tree decomposition are given in Section 3. We then explain how this framework was applied to incorporate the additional knowledge sources of accent, gender, and wide-phonetic information in Section 4. After that, we clarify how the proposed model was used in an ASR system in Section 5. Details on the experiments are then presented in Section 6, including the results and discussion. Conclusions are drawn in Section 7.

2. GENERAL FRAMEWORK FOR INTEGRATING KNOWLEDGE SOURCES

Let us first define some notations related to statistical acoustic models. We denote an HMM phonetic model by λ and $X_s = X_t, ..., X_{t+s}$ is an observation data segment of length s. Then, assume that we incorporate additional knowledge sources $K_1, K_2, ..., K_N$ into the HMM model, λ, with observation data segment X_s.

Since the knowledge sources, $K_1, K_2, ..., K_N$, might come from different domains, it may be difficult to formulate a probabilistic function of the model without learning the causal dependencies between the sources. Here, we use Bayesian network (BN) to described the causal relationship between λ, X_s, and $K_1, K_2, ..., K_N$, as that in Fig. 2, where we assume that X_s is a continuous variable denoted by a circle node, λ, $K_1, K_2, ..., K_N$ are discrete variables denoted by square nodes, and all $K_1, K_2, ..., K_N$ are conditionally independent given λ.

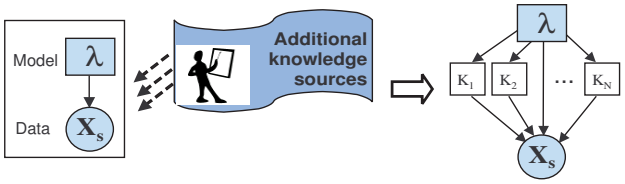

Figure 2: Describes the conditional relationship between λ, X_s, and additional knowledge sources $K_1, K_2, ..., K_N$ by using BN.

The BN joint probability function can be factorized [9] as

$$P(Z_1, Z_2, ..., Z_K) = \prod_{k=1}^{K} P(Z_k | Pa(Z_k)), \quad (1)$$

where $Pa(Z_k)$ denotes the parents of BN variables Z_k, so that we obtain

$$\begin{aligned} &P(X_s, K_1, K_2, ..., K_N, \lambda) \\ &= P(X_s | K_1, ..., K_N, \lambda) P(K_1 | \lambda) ... P(K_N | \lambda) P(\lambda), \end{aligned} \quad (2)$$

from Fig. 2. Our primary interest is to calculate probability, $P(X_s | K_1, K_2, ..., K_N, \lambda)$, which predicts data that can be expected given current knowledge about the model. Depending on the complexity of form $P(X_s | K_1, K_2, ..., K_N, \lambda)$, inference computation can be easy or difficult. If the form of this PDF is simply a single Gaussian distribution where all variables can be observed, we can simply calculate the output probability directly as

$$\begin{aligned} &p(x_s | k_{1_j}, ..., k_{N_j}, \lambda) \\ &= P(X_s = x_s | K_1 = k_{1_j}, ..., K_N = k_{N_j}, \lambda)). \end{aligned} \quad (3)$$

However, in our case, the conditional PDF involves HMM model λ and segment X_s of variable duration. Thus, the calculation of global conditional probability $P(X_s | K_1, ..., K_N, \lambda)$ might not be trivial, due to too many variables and/or computational complexity. To solve this problem, BN directed graphs need to be decomposed into clusters of variables, on which the relevant computations can be performed. This can be done with the junction tree algorithm [9], which is briefly described in the next section.

3. JUNCTION TREE DECOMPOSITION

Let us explain a simple case where we only incorporate two additional knowledge sources, K_1 and K_2. The causal relationship between X_s, λ, K_1, and K_2 is described by the BN in Fig. 3(a). Here, λ, K_1, and K_2 are discrete variables denoted by square nodes, and X_s is a continuous variable denoted by a circle node.

The following graphical transformations are then applied to obtain a junction tree [9, 10]:

1. Construct an undirected graph from the BN, by marrying the parents (adding a link between any pair of variables with a common child) and dropping the direction of the links. The resulting graph is called a **moral graph**.
2. Selectively add arcs to the moral graph to form a **triangulated graph** (adding links until all cycles consisting of more than three links have a chord).
3. Form a subset containing $Pa(A) \bigcup A$, which is called a **cluster/clique**, for all variables A with $Pa(A) \neq 0$ in the triangulated graph.
4. Build a **junction tree**, starting with clusters as the nodes, in which all links between two clusters are labeled by using a **separator** of a non-empty intersection between these clusters.

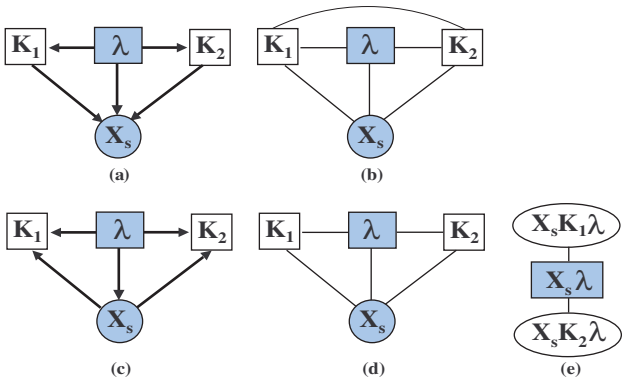

Figure 3: (a) BN topology describing conditional relationship between X_s, λ, K_1, and K_2. (b) Moral and triangulated graph of Fig. 3a. (c) Equivalent BN topology (d) Moral and triangulated graph of Fig. 3c (e) Junction tree of Fig. 3d.

Figure 3(b) outlines a moral and triangulated version of the BN from Fig. 3(a). However, we can only obtain one cluster with the full set of variables $\{X_s, \lambda, K_1, \text{and } K_2\}$ from this triangulated graph, and can not decompose any further. Fortunately, since K_1, and K_2 are assumed to be independent, by reversing some arrows we can obtain the equivalent graph as in Fig. 3(c). Figure 3(d) outlines the moral and triangulated version of this graph. We can then identify the clusters and obtain the junction tree in Fig. 3(e), where the cluster

sets are represented by oval nodes and the separator sets are represented by square nodes.

The joint probability distribution is then defined as the product of all cluster potentials, divided by the product of the separator potentials [10] as

$$P(U) = \frac{\prod_i \phi_{C_i}}{\prod_j \phi_{S_i}}, \tag{4}$$

where U is the "universe" representing all variables in the graph, ϕ_{C_i} is the cluster potential (the probability over cluster C_i), and ϕ_{S_i} is the separator potential (the probability over separator S_i). Thus, according to Fig. 3(e), joint probability function $P(X_s, K_1, K_2, \lambda)$ becomes

$$P(X_s, K_1, K_2, \lambda) = \frac{P(X_s, K_1, \lambda)P(X_s, K_2, \lambda)}{P(X_s, \lambda)}, \tag{5}$$

where $P(X_s, K_1, \lambda)$ and $P(X_s, K_2, \lambda)$ are the cluster potentials and $P(X_s, \lambda)$ is the separator potential.

Since our primary interest is to calculate $P(X_s|K_1, K_2, \lambda)$, we can derive this using Eqs. (2) and (5), and finally obtain

$$P(X_s|K_1, K_2, \lambda) = \frac{P(X_s|K_1, \lambda)P(X_s|K_2, \lambda)}{P(X_s|\lambda)}. \tag{6}$$

This demonstrates a new way of representing probability function $P(X_s|K_1, K_2, \lambda)$, as the composition of several local probability functions $P(X_s|K_1, \lambda), P(X_s|K_2, \lambda)$, corresponding to the probability of the observation data, X_s, given specific additional knowledge K_1 and K_2. In this case, the term, $P(X_s|\lambda)$, serves as a normalization constant.

Now, it should be much easier to define, estimate, and calculate several simple $P(X_s|K_i, \lambda)$ than a single but complex $P(X_s|K_1, ..., K_N, \lambda)$.

4. INCORPORATING ACCENT, GENDER, AND WIDE-PHONETIC CONTEXT INFORMATION

We first apply the approach described in the previous section to the task of incorporating additional wide-phonetic context knowledge, where K_1 is preceding context C_L and K_2 is succeeding contexts C_R. If we assume that λ is monophone unit model $/a/$, and C_L and C_R are the preceding and following context unit models $/a^-/$ and $/a^+/$, we can define the following equation

$$P(X_s|C_L, C_R, \lambda) = P(X_s|[a^-, a, a^+]), \tag{7}$$

and Eq. (6) becomes

$$P(X_s|[a^-, a, a^+]) = \frac{P(X_s|[a^-, a])P(X_s|[a, a^+])}{P(X_s|[a])}. \tag{8}$$

This equation has the same factorization as the one proposed in [11], where a triphone model is constructed from monophone and biphone models based on Bayes rule and is known as a Bayesian triphone. However, difficulties arise when different types of knowledge sources need to be incorporated.

In contrast, the current unified framework gives us a more appropriate means of incorporating various kinds of knowledge sources, not only knowledge about a wider phonetic context, but also other additional knowledge variables, such as gender (G) or accent (A) information.

One simple way of representing the composition of a pentaphone, $/a^{--}, a^-, a, a^+, a^{++}/$ is by setting λ to represent a monophone, $/a/$, and the second preceding and succeeding contexts, C_L and C_R, to represent $/a^{--}, a^-/$ and $/a^+, a^{++}/$, respectively. Then

$$P(X_s|[a^{--}, a^-, a, a^+, a^{++}]) = \frac{P(X_s|[a^{--}, a^-, a])P(X_s|[a, a^+, a^{++}])}{P(X_s|[a])}, \tag{9}$$

which indicates that pentaphone $P(X_s|[a^{--}, a^-, a, a^+, a^{++}])$ can be composed from left/preceding-triphone-context unit (L3), right/following-triphone-context unit (R3), and monophone unit (C1). We call this composition C1L3R3.

Then, we can further extend C1L3R3 with gender (G) and accent (A) information. We call this composition C1L3R3-AG. The likelihood function is obtained using the same consideration and is expressed as

$$P(X_s|[a^{--}, a^-, a, a^+, a^{++}], A, G) = \frac{P(X_s|[a^{--}, a^-, a], A, G)P(X_s|[a, a^+, a^{++}], A, G)}{P(X_s|[a], A, G)}, \tag{10}$$

which indicates that $P(X_s|[a^{--}, a^-, a, a^+, a^{++}], A, G)$ can be calculated by factorizing probabilities $P(X_s|[a], A, G)$, $P(X_s|[a^{--}, a^-, a], A, G)$, and $P(X_s|[a, a^+, a^{++}], A, G)$.

5. USE OF PROPOSED MODEL

We used the proposed models by rescoring the N-best list generated from a standard and unmodified triphone ASR system to avoid decoding complexity, as summarized in Fig. 4.

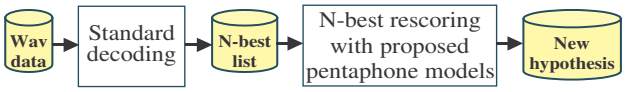

Figure 4: Rescoring procedure with proposed models.

N-best recognition (on the word level) was carried out for all utterances in the test data using a conventional HMM model and standard two-pass decoding based on a Viterbi algorithm. Each N-best hypothesis included an acoustic score, a language modeling (LM) score, and a Viterbi segmentation of all phonemes. Each phoneme segment in each hypothesis was then rescored using the pentaphone C1L3R3 models, summarized in Fig. 5. These updated acoustic scores were combined with the LM score for this hypothesis. The hypothesis achieving the highest total utterance score of the N-best hypothesis was eventually selected as the new recognition output.

The parameter estimation of the proposed pentaphone model may become unreliable, if there are insufficient training data, as will state output. We used deleted interpolation to improve reliability, which allowed us to fall back to a more reliable model when the supposedly more precise model was, in fact, unreliable [12]. The concept usually involves interpolating two (or more) separately trained models, one of which is more reliably trained than the other. Instead of interpolating two models, we applied this approach to interpolating two phonetic likelihoods, where the phonetic likelihood of the proposed pentaphone model, $P(X_s|\lambda_{pentaphn})$, was the more precise, while the triphone likelihood, $P(X_s|\lambda_{triphn})$,

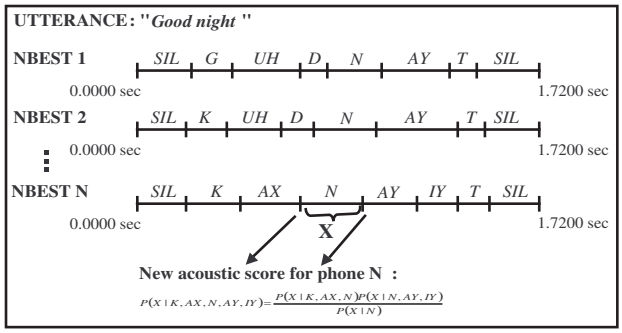

Figure 5: N-best rescoring mechanism.

was the more reliable. Consequently, the interpolation phonetic likelihood, $P(X_s|\lambda)$, is obtained as

$$P(X_s|\lambda) = \alpha P(X_s|\lambda_{pentaphn}) + (1-\alpha)P(X_s|\lambda_{triphn}), \quad (11)$$

where α represents the weight of the HMM phonetic likelihood of the proposed pentaphone model, and $(1-\alpha)$ represents the weight of the HMM phonetic likelihood of the triphone model. If the amount of training data is sufficiently large, $P(X_s|\lambda_{pentaphn})$ becomes more reliable and α is expected to tend to 1.0. However, if it is not, α will tend to 0.0 so as to fall back to the more reliable model, $P(X_s|\lambda_{triphn})$.

6. EXPERIMENTS

The experiments were conducted using feature extraction parameters that were a sampling frequency of 16 kHz, a frame length of a 20 ms Hamming window, a frame shift of 10 ms, and 25-dimensional feature parameters consisting of 12-order MFCC, Δ MFCC and Δ log power. The speech corpus used here was an accented English speech corpus based on travel domain expressions. It consisted of American (US) and Australian (AUS) English accents, with about 45k utterances (44 speech hours) spoken by 100 speakers (50 males, and 50 females) for each accent. About 40k utterances (90% of the data) spoken by 80 speakers (40 males, and 40 females) was used as the training data. Two hundred utterances randomly selected from the remaining 10% were used as the test data. We used both bi-gram and tri-gram language models, which were trained on about 150,000 travel-related sentences. The available pronunciation dictionary consisted of about 37k words and was based on US pronunciations.

Three states were used as the initial HMM for each phoneme. A shared state HMnet topology was then obtained using a successive state splitting (SSS) training algorithm. Since the SSS algorithm used was based on the minimum description length (MDL) optimization criterion, the number of shared HMM states was determined automatically by the algorithm. Details on MDL-SSS can be found in [13]. A context-dependent triphone system having 2,126 total states with four different versions of Gaussian mixture components per state, i.e., 5, 10, 15, and 20, was used as the baseline. Additional knowledge such as gender and accent can also be incorporated in the conventional triphone acoustic model (AM) by training gender and/or accent dependent AMs. Only an embedded training procedure was conducted with specific accent or gender training data to create the same topology structure for all models. Thus, in total, we obtained one single triphone AM (without any additional knowledge) and four accent-gender-dependent triphone AMs (for US males and females, and AUS males and females).

Each component of the C1L3R3 model was trained separately using the same amount of training data and the same SSS training algorithm. There was a total of 3,403 states (sum of C1: 132 st., L3: 1,645 st., R3: 1,626 st.) and the same number of Gaussian mixture components as the baseline. An embedded training procedure was then carried out on C1L3R3-AG on specific accent or gender training data.

We first evaluated the advantages of incorporating additional knowledge sources in multi-accented test data. Rescoring was done using a 10-best list, and a 0.3 weight parameter, α, for deleted interpolation was used, which had been optimized using a development set as in our previous study [14]. We also conducted additional experiments with a conventional pentaphone HMM model with 2,202 states, which was trained from scratch using MDL-SSS, for comparison. Accent- and gender-dependent pentaphone models were also obtained using an embedded training procedure on all specific accents or gender training data. They were implemented by rescoring the N-best list as in the case of the proposed pentaphone C1L3R3.

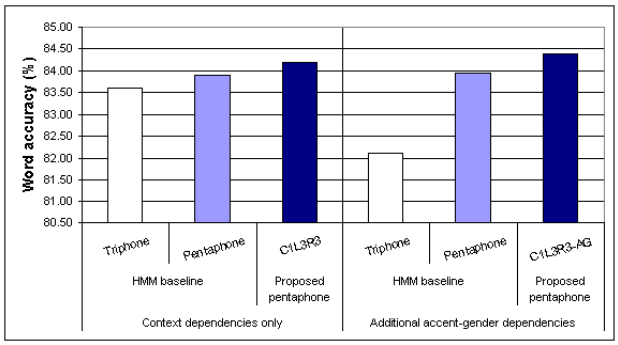

Figure 6: Comparison of recognition accuracy rates of different systems triphone HMM baselines, pentaphone HMM baselines, and proposed pentaphone models.

The performance of models having five mixture components per state is depicted in the bar graph in Fig. 6. The triphone baseline without any additional knowledge achieved 83.60% word accuracy. However, this decreased to 82.11% word accuracy for accent-gender-dependent models. This might be due to the size of the training data which is much smaller than that for the other baseline models. Performance could improve up to 83.96% word accuracy by rescoring with more precise model such the conventional pentaphone HMM. There was no decrease performance when gender and accent were incorporated, as in the case of triphone baseline, which is probably due to the use of deleted interpolation. However, as can be seen, the proposed pentaphone C1L3R3 model could considerably outperform the conventional pentaphone HMM. This might be because, given the amount of training data, training the conventional pentaphone model using the MDL-SSS algorithm resulted in a model with 2,202 total states, which is not so different from the total number of states in the triphone HMM. As many different pentaphone contexts may have shared the same Gaussian components, the context resolution was reduced. Thus, approximating a pentaphone model using a composition of several less context-dependent C1L3R3 models could help to reduce the

loss of context resolution and improve performance. Overall, the best word accuracy that was achieved was 84.38% with C1L3R3-AG, which incorporated additional knowledge of accent A, gender G, second preceding context C_L and succeeding context C_R.

We next investigated the improvements in performance in more detail using pentaphone C1L3R3-AG on each accented test data, with the N-best (N=10) list. We used the 0.3 weight parameter λ for deleted interpolation as in our previous study [14]. Here, we also measured both the relative improvement (Rel-Imp) and a relative rescoring improvement (Rel-Resc-Imp) as used in [6]

$$\text{RelRescImp} = \frac{\text{Rescoring result} - \text{Baseline}}{\text{Nbest list upper bound} - \text{Baseline}}, \quad (12)$$

where the N-best list upper bound is the N-best recognition result when the best match candidates are chosen.

The results obtained using the different mixture component numbers are summarized in Table 1 for US test data and Table 2 for AUS test data. As can be seen, the proposed pentaphone model consistently improved the performance of the ASR system. The largest Rel-Resc-Imp achieved 37.92% for the US model with 15 mixture components per state, along with 38.04% for the AUS model with 15 mixture components per state.

Table 1: Recognition accuracy rates for pentaphone C1L3R3-AG on US test data.

# Mix	Upper bound	Triphn baseline	Proposed pentaphn	*Rel-Imp*	*Rel-Resc-Imp*
5	87.52	84.30	85.19	*5.67*	*27.64*
10	87.94	84.66	85.79	*7.37*	*34.45*
15	87.76	84.78	85.91	*7.42*	*37.92*
20	87.78	85.25	85.91	*4.47*	*26.09*

Table 2: Recognition accuracy rates for pentaphone C1L3R3-AG on AUS test data.

# Mix	Upper bound	Triphn baseline	Proposed Pentaphn	*Rel-Imp*	*Rel-Resc-Imp*
5	85.79	82.33	83.76	*8.09*	*41.33*
10	85.37	82.21	82.81	*3.37*	*18.99*
15	86.93	83.46	84.78	*7.98*	*38.04*
20	86.39	82.63	83.58	*5.47*	*25.27*

7. CONCLUSION

We introduced a general framework to incorporate additional knowledge sources into statistical HMM acoustic models. We also demonstrated the implementation of this new framework by integrating accent, gender, and wide-phonetic context information. The framework is based on a junction tree algorithm and allows us to construct wider context models from several other models with a narrower context. As this leads to a reduction in the number of context units to be estimated, the loss of context resolution can be considerably reduced. We applied these composition models at the post-processing stage with N-best rescoring. Their performance was evaluated on an LVCSR task using two different types of accented English speech data. Experimental results demonstrated that our method improved word accuracy with respect to the standard HMM with or without additional knowledge sources. The best performance was obtained by a model that incorporated additional knowledge on accent A, gender G, second preceding context C_L and succeeding context C_R.

REFERENCES

[1] W.J. Holmes and M. Huckvale, "Why have hmms been so successful for automatic speech recognition and how might they be improved," *Speech, Hearing and Language*, vol. 8, pp. 207–219, 1994.

[2] D.H. Klatt, "Review of the ARPA speech understanding project," *Acoustical Society of America*, vol. 62, pp. 1345–1366, 1977.

[3] V.W. Zue and R.A. Cole, "Experiments on spectrogram reading," in *Proc. ICASSP*, Washingtton D.C., USA, 1979, pp. 116–119.

[4] J. Johannsen, J. MacAllister, T. Michalek, and S. Ross, "A speech spectrogram expert," in *Proc. ICASSP*, Boston, USA, 1983, pp. 746–749.

[5] S.E. Levinson, "Structural methods in automatic speech recognition," in *Proc. IEEE*, November 1985, vol. 73, pp. 1625–1650.

[6] J. Li, Y. Tsao, and C.-H. Lee, "A study on knowledge source integration for candidate rescoring in automatic speech recognition," in *Proc. ICASSP*, Philadelphia, USA, 2005, pp. 837–840.

[7] G. Zweig and S. Russell, "Probabilistic modeling with Bayesian networks for automatic speech recognition," in *Proc. ICSLP*, Sydney, Australia, 1998, pp. 3010–3013.

[8] K. Daoudi, D. Fohr, and C. Antoine, "A new approach for multi-band speech recognition based on probabilistic graphical models," in *Proc. ICSLP*, Beijing, China, 2000, pp. 329–332.

[9] F. Jensen, *An Introduction to Bayesian Network*, UCL Press, 1998.

[10] C. Huang and A. Darwiche, "Inference in belief networks: A procedural guide.," *International Journal of Approximate Reasoning*, vol. 11, pp. 1–158, 1994.

[11] Ji Ming, P. O Boyle, M. Owens, and F. Jack Smith, "A Bayesian approach for building triphone models for continuous speech recognition," *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 6, pp. 678–684, November 1999.

[12] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing*, Prentice Hall, New Jersay, USA, 2001.

[13] T. Jitsuhiro, T. Matsui, and S. Nakamura, "Automatic generation of non-uniform HMM topologies based on the MDL criterion," *IEICE Trans. Inf. & Syst.*, vol. E87-D, no. 8, pp. 2121–2129, 2004.

[14] S. Sakti, S. Nakamura, and K. Markov, "Improving acoustic model precision by incorporating a wide phonetic context based on a Bayesian framework," *IEICE Trans. Inf. & Syst.*, vol. E89-D, no. 3, pp. 946–953, 2006.