Limited Distribution

TR-I-0115

Research Activities of the Speech Processing Department

January 1989 through October 1989

ATR Interpreting Telephony Research Laboratories

November, 1989

ATR Interpreting Telephony Research Laboratories

Research Activities of
the Speech Processing Department

January 1989 through October 1989

Abstract:

The research areas of the Speech Processing Department include:

(A) Large Vocabulary Continuous Speech Recognition.
- (1) Feature-Based Approaches.
- (2) Hidden Markov Models.
- (3) Neural Network Approaches.

(B) Speaker Adaptation.
- (1) Speaker Adaptation by Codebook Mapping.

(C) Language Source Modeling.
- (1) Language Source Modeling by Generalized LR Parser.
- (2) N-Gram Word Prediction by Neural Networks.
- (3) Phoneme Source Modeling.

(D) Noise Reduction.
- (1) Noise Reduction by Neural Networks.
- (2) Noise Tolerant Speech Recognition Using Codebook Mapping.

(E) Speech Synthesis by Rule.
- (1) High Quality Japanese Speech Synthesis by Rule.
- (2) Prosodies and Speaking Styles.

(F) Voice Conversion.
- (1) Voice Conversion by Codebook Mapping.
- (2) Speech Analysis and Synthesis Algorithm.

(G) Phonetics
- (1) Analysis of Allophones and Coarticulation.

(H) Speech Database
- (1) Large-Scale Speech Database with Phonetic Transcriptions.
- (2) Speech Database Retrieval System.
- (3) Speech Workbench.

For additional information, contact:

ATR Interpreting Telephony Research Laboratories

Inuidani, Seika-cho, Souraku-gun, Kyoto 619-02, JAPAN

Telephone: +81−7749−5−1311, Telefax: +81−7749−5−1308

E-mail: atr-la.atr.jp ! shikano @ uunet.UU.NET (from USA arpanet).

Contents

1. General

The necessity for automatic telephone interpretation research is described in brief. We introduce the Speech Processing Department of the ATR Interpreting Telephony Research Laboratories, and provide an overview of its research activities from January 1989 through October 1989.

1.1. Automatic Telephone Interpretation Project

The Automatic Telephone Interpretation system is a facility which enables a person speaking in one language to communicate readily by telephone with someone speaking another language. It does so by automatically and simultaneously transforming the dialogue from the speaker's to the listener's language. At least three constituent technologies are necessary for such a system: speech recognition, machine translation and speech synthesis.

Since this is a new concept, a number of feasibility studies and evaluations must still be made. A high degree of performance from each of the constituent technologies, and the integration of these technologies and system user friendliness are also essential. According to a feasibility study report published in 1985 by the Japanese Ministry of Posts and Telecommunications, realizing a system like this will require at least fifteen years. The research in this project also includes various kinds of multi-media technologies such as speech recognition, machine translation and speech synthesis. These technologies will be applied to a multi-media workstation environment not only to allow people speaking the same language to communicate with each other but also people speaking different language.

To do so, one more important requirement of this ambitious project is the ongoing cooperation of research institutions in Japan and abroad.

1.2. Speech Processing Department

The ATR Interpreting Telephony Research Laboratories, Dr. Akira Kurematsu, president, is made up of three departments. They are the Natural Language Understanding Department, the Knowledge and Database Department and the Speech Processing Department. At least three basic technologies, including speech recognition, machine translation and speech synthesis, are necessary for such an interpreting telephony system. Moreover, integrated research into these technologies is also very important. We propose the interpreting telephony model shown in Figure 1-1. In this model, language processing is split into a language source modeling stage and a language analysis stage. The main targets of our research laboratories are fundamental research into speech and language processing, and integration of speech and language processing technologies to demonstrate the feasibility of an automatic telephone interpretation system.

These laboratories were founded in April of 1986 with the support of the Japan Key Technology Center, ATR International, NTT, KDD, NHK and other Japanese enterprises.

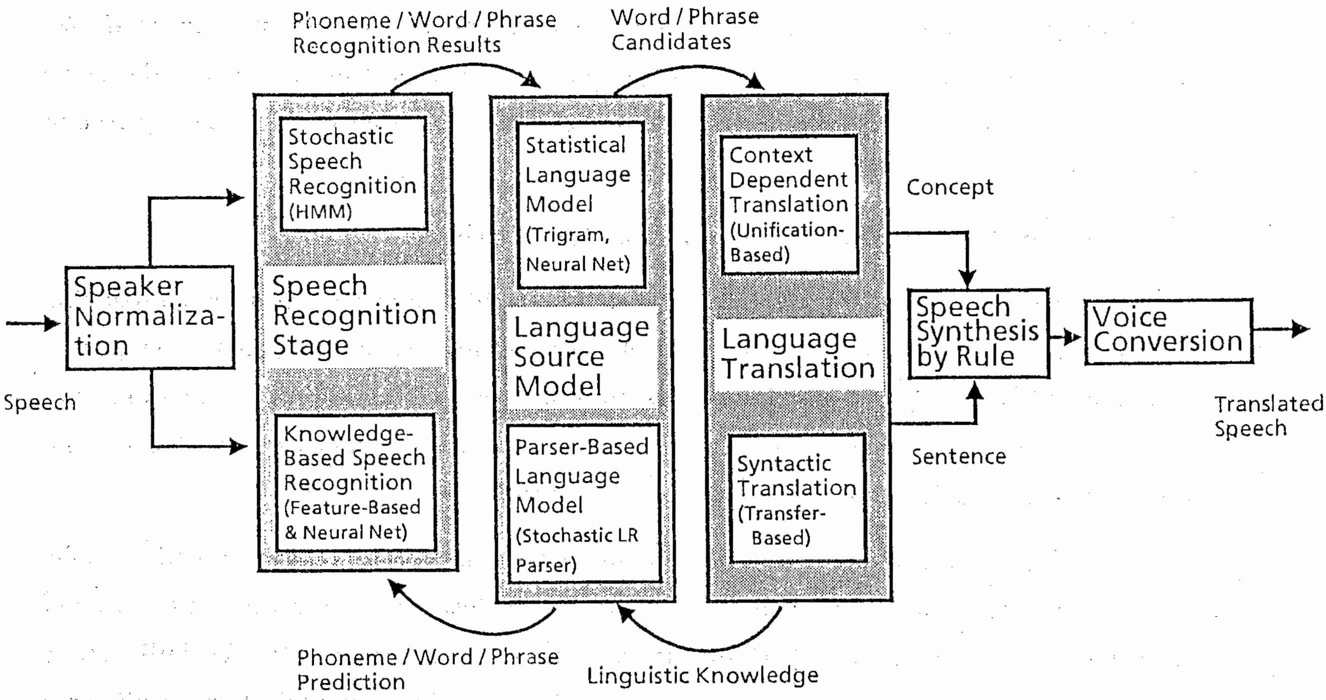

Figure 1-1. Proposed Interpreting Telephony Experimental System.

The Speech Processing Department, headed by Dr. Kiyohiro Shikano, is one part of this endeavor. The research areas of the Speech Processing Department are speech recognition , speech synthesis and speech database. These research areas are aimed at advanced technologies capable of recognizing continuous speech from any speaker and synthesizing high quality speech with the characteristics of various speakers. These technologies, which are indispensable to realize an interpreting telephony system, must be considerably improved by investigating these fields and exploring innovative techniques and break-through technologies.

The Speech Processing Department is comprised of two major groups: a speech recognition group and a speech synthesis and phonetics group. The speech recognition group is mainly directed by Drs. K. Shikano, T. Kawabata and H.Sawai. The speech synthesis and phonetics group is mainly directed by Drs. Y. Sagisaka, T. Umeda and S.Tamura.

As of October 1989, efforts aimed at speaker-dependent phoneme recognition and speaker-independent phoneme segmentation have resulted in dramatically improved phoneme recognition and continuous speech recognition performance. A continuous speech recognition system based on Hidden Markov Modeling attains a phrase recognition rate of 88% for a 1,000 word task. We have just started to implement continuous speech recognition hardware.

We are taking three approaches to speech recognition: (1) *The Feature-Based* approach, especially for phoneme segmentation, (2) *The Hidden Markov Model* approach, and (3) *The Neural Network* approach. Continuous speech recognition and word spotting efforts have also continued. For speaker-independent speech recognition, a speaker adaptation approach has been undertaken using the concept of *Fuzzy Vector Quantization and Spectrum Mapping*. We also successfully applied this algorithm to an HMM-based continuous speech recognition system with a phrase recognition rate of 79%. In the area of language models, the *N-Gram* approach using Neural Networks and a language source modeling approach using the *generalized LR* parser have been followed.

Analysis efforts for the exploration of *Allophones and Coarticulation* phenomena have been undertaken utilizing a large-scale speech database. In order to realize a high quality rule-based speech synthesis system, we are working in several areas, such as a speech synthesis by rule based on *Multiple Speech Units* from a phoneme-labelled speech database, high quality speech analysis and synthesis algorithms, and voice conversion algorithms from one speaker to another based on *Vector Quantization and Mapping (Codebook Mapping)*. Moreover, non-parametric noise reduction algorithms using *Neural Networks* have been investigated, and cross-linguistic voice conversion from MITALK to a Japanese speaker has been carried out successfully.

A large-scale speech database with phonetic transcriptions, mainly for research in speech recognition and speech synthesis, has been developed with considerable support from other speech research laboratories. A CD ROM of the speech database will be available for public use shortly. A speech workbench and a database retrieval system have been implemented on a workstation in order to provide researchers with a tool for easy access to, and manipulation of, speech data.

2. Research Activities

Research activities in the Speech Processing Department from January 1989 through October 1989 are summarized below, and related technical publications are quoted.

2.1. Large Vocabulary Continuous Speech Recognition

Reliable phoneme recognition and segmentation algorithms have been investigated leading to considerable improvements over conventional approaches. We have been pursuing three approaches: a *Feature-Based* approach, a *Hidden Markov Model* approach, and a *Neural Network* approach. These improvements resulted in the successful implementation of a continuous speech recognition system such as *HMM-LR* by combining HMM phoneme models with the generalized LR parsing algorithm. Hardware implementation has also begun.

2.1.1. Feature-Based Approaches [Komori, Hatazaki, Tanaka, Kitagaito, Kawabata, Shikano]

A spectrogram reading seminar taught by Prof. Victor Zue was held in February of 1986. Since that time, the spectrogram reading session had been convened on a weekly basis. Many researchers and phoneme labelers in our department can now read phonemes of Japanese utterances with relatively high accuracy.

In particular, phoneme boundaries can be identified correctly in Japanese utterances using expert knowledge, which can deal with various kinds of coarticulation phenomena. Now we are developing an expert system for phoneme segmentation using the expert tool, *ART*. A Symbolics lisp machine for running *ART* and a micro Vax station/Vax 8800 for feature extraction of input speech are connected and used to implement a phoneme segmentation expert system. The phoneme segmentation knowledge for consonants is described by ART. The phoneme segmentation expert is integrated with a *TDNN* neural network for consonant discrimination. The system attains a consonant segmentation rate of 94.5% and a consonant recognition rate of 88.8%. Moreover, spotting of vowels, semi-vowels and the syllabic nasal has been studied using a vowel spotting TDNN neural network.

The final aim of this feature-based approach is the implementation of a phonemic typewriter, which is able to recognize continuous speech without language knowledge such as words or syntax.

Technical Publications: [Hatazaki-89-05, Komori-89-06, 09], [The phoneme segmentation expert integrated with a *TDNN* neural networks will be presented at ICASSP90, and related technical reports in Japanese are also available]

2.1.2. Hidden Markov Models [Hanazawa, Kawabata, Maruyama, Kita, Shikano]

Hidden Markov Models (HMM) were studied to determine how to best use models in the training stage on the forward-backward algorithm. The number of states, tied probabilities of transition and output, probability smoothing techniques, initial probability settings and duration control techniques were studied based on the task of phoneme recognition using our large vocabulary speech database described in section **2.8.1**. After that the HMM phoneme models were improved by introducing *Separate VQ* (multiple codebooks) and *Fuzzy VQ* techniques. These improvements resulted in a 9% phoneme recognition rate improvement, from 86.5% to 95.7%.

The Hidden Markov models based on phoneme units have been applied successfully to word spotting in continuous speech. Moreover, the HMM phoneme models are combined with a generalized LR parser (a language source model) to efficiently recognize a Japanese phrase input, where the LR parser can predict phoneme candidates to the HMM phoneme models. This system is called *HMM-LR*. HMM-LR can recognize Japanese phrase inputs with a phrase recognition accuracy of 88%. The performance correct for the top-two and top-five phrase candidates are 96% and 99%, respectively. Robustness studies for continuous speech recognition and speaker adaptation have been initiated to improve HMM-LR performance. HMM-LR hardware implementation has also been initiated in order to show the feasibility of a prototype interpreting telephony system. The HMM phoneme model approach is also applied to English word recognition.

Technical Publications: [Hanazawa-89-10a, 10b, Kawabata-89-05, Kita-89-05, 06, 08, Maruyama-89-01, 10], [Stochastic HMM-LR and the HMM-LR performance will be presented at ICASSP90]

2.1.3. Neural Network Approaches [Sawai, Minami, Miyatake, Dang, Tamura, M.Nakamura, Kawabata, Waibel, Shikano]

Using Neural Networks, phoneme recognition and continuous speech recognition have been investigated. A time-delay neural network (*TDNN*) now achieves high phoneme recognition rates for the task of speaker-dependent discrimination not only among the voiced consonants, /b/,/d/, and /g/ but also for full consonants. All consonants were extracted from the phonetically labelled large vocabulary database , i.e., from 5,240 common Japanese words spoken by three speakers. The TDNN attains a 98.6% phoneme recognition rate for the /b/,/d/, and /g/ task, and 96.7% for the full consonant task. Phoneme spotting by TDNN aimed at determining a continuous speech recognition approach by neural networks has been carried out and shows a phoneme spotting rate of 98%. A preliminary experiment to recognize continuous utterances using TDNN phoneme spotting results has been tried by summing up the phoneme spotting outputs by means of a DTW algorithm.

Efforts to speed up the back-propagation algorithm resulted in a 1,000-fold speedup of the TDNN. This speedup and the use of an Alliant mini-supercomputer make it possible to challenge larger scale neural networks. TDNN robustness and generalization studies for different speaking styles have been initiated.

Other neural network approaches such as a deterministic Boltzmann machine and a neural prediction model are also being studied. A constructive neural network (CNN) to recognize words in bottom-up fashion using a TDNN phoneme spotting network is also studied.

Technical Publications: [Dang-89-10, Kawabata-89-10, Minami-89-10, Miyatake-89-06, Sawai-89-03, 05, 06, 08, Waibel-89-03, 05], [Phoneme spotting based on TDNN will be presented at ICASSP90, and related ATR technical reports are also available on request]

2.2. Speaker Adaptation

Aiming at a general preprocessor for speaker normalization, speaker adaptation research using *codebook mapping* techniques has been studied. Discrete spectrum space representation by vector quantization makes it possible to realize sophisticated speaker adaptation / normalization by codebook mapping. In paticular, the fine tuning to HMM has been investigated.

2.2.1. Speaker Adaptation by Vector Quantization[S.Nakamura, Hattori, Shikano]

As a general preprocessor to a phoneme recognizer, speaker normalization algorithms have been developed. The algorithms adopt vector quantization as a discrete representation of spectral space. Discrete representation makes it possible to carry out sophisticated spectral normalization or mapping from one speaker to another. In previous studies, algorithms using a single codebook were developed, and a complex spectral distortion measure which is composed of a spectrum term, a differenced spectrum term and a differenced power term was adopted. The complex spectrum distortion measure improved the word recognition results, but also significantly increased spectrogram distortion. In order to reduce this degradation, algorithms with multiple codebooks (separate VQ) were investigated. *Fuzzy vector quantization* techniques have also been investigated to realize more accurate speaker adaptation. These algorithms are also successfully applied to voice conversion, as described in section **2.6**. These algorithms are also applied to neural network speech recognition and will be applied to feature-based speech recognition as a speaker normalization preprocessor.

These technologies lead us to an HMM speaker adaptation algorithm greatly improved over previously reported algorithms. The HMM speaker adaptation algorithm attains a phrase recognition rate of 78.6%, where the phrase recognition rates of speaker independence and speaker dependence are 59.6% and 88.4%, respectively. The correct rates for the top-two and top-five phrase candidates are 92% and 97%, respectively. Moreover, a supplement HMM phoneme model training approach has been initiated to cope with speaker coarticulation variations.

A comparative study between nonlinear neural network mapping and codebook mapping is performed, which clarify that codebook mapping is more accurate than neural network mapping.

Technical Publications: [S.Nakamura-89-02, 05, 06, 10a, 10b], [Speaker adaptation research progress will be presented at ICASSP90, and related technical reports are also available on request]

2.3. Language Source Modeling

The ATR Interpreting Telephony Laboratories also include two other departments which are studying language models more deeply than the Speech Processing Department. Nevertheless, the Speech Processing Department has been studying language source models based on bottom-up word/phoneme prediction and statistical word/phoneme prediction in collaboration with the other departments. We have been taking two approaches: an LR source modeling approach and a neural network approach (*Netgram* source modeling), to predict the next word sets or phonemes. Such bottom-up word/phoneme prediction approaches based on language source modeling should be combined with top-down approaches in deeper language processing.

Another approach to making a phoneme sequence source model has been initiated aiming at a phonemic typewriter.

2.3.1. Language Source Modeling by Generalized LR Parser [Kita, Kawabata, Hanazawa, Shikano]

Trigram language models based on word categories were introduced in order to improve word recognition results for English sentences uttered word by word. Probabilities of the trigrams of word categories were estimated using the Brown Corpus Text Database. Moreover, high probabilities of common word sequences such as frozen word sequences were extracted from the Brown Corpus Text Database. These probabilities were integrated using a dynamic programming word recognition algorithm, and improved word recognition from 80.9% to 89.1%.

The generalized LR parser, called the LR parser for short, was introduced to predict next words/phonemes. The LR parser can be regarded not only as a parser for input word sequences but also a language source model for word/phoneme prediction/generation. The LR parser can deal with a context-free grammar. The LR parser is successfully integrated with HMM phoneme models to recognize a Japanese phrase utterance, where the LR parser very quickly predicts phoneme candidates according to the LR table. The LR parser will be developed for better language source modeling by introducing the probabilities of phoneme sequences and word sequences automatically estimated from the large-scale text database. The combined system of the LR language source model and HMM, called *HMM-LR,* attains a phrase recognition rate of 88% for a task with a phoneme perplexity of 5.9. The task includes 1,000 words and is written using a context-free grammar to deal with Japanese phrases(Bunsetsu). The cooccurrence of context-free rewriting rules will be used to produce a better language source model.

The LR parser is also successfully combined with the Sphinx system at Carnegie Mellon University(CMU) under the research collaboration between CMU and ATR.

Other low-level linguistic information is contained in phoneme sequences. The use of phoneme sequence information such as a syllable trigram has been studied aiming at source modeling of Japanese speech inputs.

Technical Publications: [Kita-89-05, 08, 10], [The stochastic LR source modelling will be presented at ICASSP90]

2.3.2. N-Gram Word Prediction by Neural Networks [M. Nakamura, Maruyama, Shikano]

A neural network approach, *the Netgram,* has been developed to predict words using the Brown Corpus Text Database. The results have been compared to the results of statistical trigram modeling. The analysis of the internal representation of the Netgram reveals some correspondence to language categories. In order to realize this approach practically and quickly, further improvements to our learning network simulator have been made. The Netgram output values are successfully used to improve word recognition in English sentences. This approach will make the realization of a phonemic typewriter more feasible.

Technical Publications: [M.Nakamura-89-05, Maruyama-89-10], [The NETgram application to English sentence recognition will be presented at ASA Fall Meeting 1989, and related ATR technical reports are available on request]

2.3.3. Phoneme Source Modeling [Kawabata, Hanazawa, Shikano]

Aiming at a phonemic typewriter of Japanese continuous speech input, phoneme source modeling has just been initiated using a Japanese text database. First, a syllable trigram is calculated from the Japanese text database. The syllable trigram phoneme source model is then used as a phoneme prediction model, and is combined with the HMM speech recognition algorithm based on phoneme models. This results in a phoneme recognition rate of 90%. More accurate phoneme source modeling is being studied using Japanese text database.

2.4. Noise Reduction

In order to bring speech recognition technologies into the field, noise reduction must be implemented. We are now looking for new noise reduction mechanisms using various kinds of neural networks. We proposed a noise reduction algorithm using a neural network whose input and output are a speech wave itself. The algorithm successfully extends neural network application to the noise reduction problem.

Another practical approach to noise tolerant speech recognition is the use of a codebook mapping algorithm, which is developed mainly for speaker adaptation. The codebook mapping algorithm will also be evaluated using noisy speech database.

2.4.1. Noise Reduction by Neural Networks [Tamura]

Noise reduction algorithms using neural networks have been under investigation using the back-propagation algorithm on multi-layered feed-forward networks. During training, the networks develop their internal representation for a mapping from a set of noisy signals to the set of noise-free signals with noise-added speech signals as input and noise-free speech signals as the target output. The training was carried out using the digitized speech wave itself. Noise reduction algorithms already effectively eliminate computer room noise. Effectiveness for other noise and other speakers was also studied. The internal representation of the noise reduction networks was investigated, and has clarified the noise reduction mechanism.

Technical Publications: [Tamura-89-01, 05, 06], [The research progress of the noise reduction neural network will be presented at ICASSP90, and related noise reduction networks are also presented at ICASSP90]

2.4.2. Noise Tolerant Speech Recognition Using Codebook Mapping

[Ohkura, Hattori, Umeda]

Another promising approach to noise tolerant speech recognition seems to be the use of the codebook mapping algorithm, which is developed for speaker adaptation as described in section 2.2. To evaluate the codebook mapping algorithm for noisy speech, we have just started to make several kinds of noisy speech databases, which include additional noises and reverberation noises. The evaluation experiments for the codebook mapping algorithm will be executed shortly.

2.5. Speech Synthesis by Rule

A high quality speech synthesis system by rule should be realized as an important part of the interpreting telephony system. A synthesis system based on flexible synthesis units has been developed. Moreover, the speech synthesis algorithm is evaluated to realize a better synthesis system. A unit selection algorithm from a large-scale speech database and a unit concatenation algorithm have been studied. The prosody control algorithm has been also studied using a pitch database.

Also, rules between speaking styles and prosodic control have been studied to realize various kinds of synthesized voices.

2.5.1. High Quality Japanese Speech Synthesis by Rule [Sagisaka, K. Abe, Takeda]

In speech synthesis by rule, various kinds of synthesis units such as phonemes, dyads, CV-syllables and CVC units have been proposed and tried. Our proposed synthesis scheme is different from ordinary systems in two ways. One is the flexible use of non-uniform synthesis units, and the other is the choice of optimal units of multi-templates using criteria such as the feasibility of unit concatenation and appropriateness of the units to the target context. The strength of coarticulation for CV syllables is also investigated in depth. From the viewpoint of coarticulation phenomena, the place of articulation is much more dominant than the manner of articulation.

A unit selection algorithm from a large-scale speech database has been studied from the point of view of unit concatenation feasibility and coarticulation naturalness. A unit concatenation algorithm has been implemented to reduce concatenation distortion.

Prosody control models to generate natural Japanese synthesized utterances are studied using the speech database with pitch information. A more detailed prosody control model has been developed with neural networks which use large pitch and text database for training.

Speech synthesis algorithms are evaluated over LPC, Cepstrum, and PSE. OLA(overlapping add) synthesis algorithms are also evaluated, and we adopted an improved cepstrum analysis/synthesis algorithm augmented by the OLA synthesis algorithm.

A prototype speech synthesis-by-rule system is implemented, and efforts in improvement and evaluation continue.

Technical Publications: [K.Abe-89-03, 10, Sagisaka-89-05, 10, Takeda-89-05, 09, 10], [Related ATR technical reports are available on request]

2.5.2. Prosodics and Speaking Styles [Miyatake, Poser, Sagisaka]

Effects of speaking styles on prosodic parameters were analyzed to clarify the inter-relationship of pitch patterns, power patterns and durations. The analysis shows that differences among speaking styles are strongly reflected in pitch and power patterns. The results suggest the possibility of a new prosody control model for speech synthesis. Prosodic models have been studied to realize more natural synthesized speech by rules.

The down-step model to deal with Japanese natural prosody has also been studied in order to realize a better prosody control model.

Technical Publications: [Poser-89-01a, 03a, 03c, Sagisaka-89-10], [Related ATR technical reports are available on request]

2.6. Voice Conversion

Voice conversion from one speaker to another is an important aspect of realizing an automatic telephone interpreting system. Voice conversion algorithms based on the codebook mapping by vector quantization have been developed. We have also started to analyze personal voice characteristics using analysis-by-synthesis methods. Moreover, *cross-linguistic voice conversion* from MITALK to a Japanese female voice is tried successfully. A new speech analysis and synthesis algorithm based on short-time Fourier transform was investigated in order to deal with more personal voice characteristics.

2.6.1. Voice Conversion by Codebook Mapping [M. Abe, Kuwabara, S.Nakamura, Umeda, Shikano]

A new voice conversion technique through vector quantization has been studied. The algorithms used in this voice conversion are basically the same as those used in the codebook mapping described in section **2.2**. The algorithms have three codebooks, one each for spectrum, power and pitch. In order to interpolate spectra from spectrum vectors in a codebook, the fuzzy vector quantization algorithm is also introduced. Voice conversion from male voice to female voice was tried and a listening test showed that the converted voice was uniformly judged to be female.

In order to improve the converted voice, fuzzy vector quantization techniques are adopted, and the vector quantization of residual waves is being explored.

Moreover, a preliminary experiment in *cross-linguistic voice conversion* from MITALK to a Japanese female voice is successfully carried out. The quality of cross-linguistic voice conversion is evaluated using the mutual information of codebook mapping probabilities. In order to improve the quality of cross-linguistic voice conversion, correspondence details of codebook mapping between Japanese and English will be studied.

Aiming at better speech quality, the modification of formant frequencies and bandwidths will also be studied. These modification techniques will be applied to analyzing personal voice characteristics. Attention is now focused on the dynamic aspects of the spectral parameters that convey personal voice characteristics.

Technical Publications: [M.Abe-89-03] , [The cross linguistic voice conversion will be presented at ICASSP90, and related technical reports are also available on request]

2.6.2. Speech Analysis and Synthesis Algorithm [M. Abe, Tamura, Umeda]

A new speech analysis and synthesis algorithm based on short-time Fourier transform has been investigated aiming at better separation between spectral envelope information and vocal source information not only for voice conversion but also for a high quality synthesis by rule system.

Technical Publications: [M.Abe-89-05, 08]

2.7. Phonetics

Analysis of speech phenomena has been started using our large vocabulary speech database in order to clarify allophonic and coarticulatory phenomena.

2.7.1. Analysis of Allophones and Coarticulation [Takeda, Kuwabara, Cheveigne, Umeda, Sagisaka]

As a first step in analyzing allophonic and coarticulatory phenomena, devocalization of vowels was studied using the large vocabulary speech database. The contexts of the devocalization are predicted by rules. Next, phoneme durations were studied using the large speech database.

Accentuation of Japanese language and problems of phonological phrase are studied using the pitch database. Voiceless long stops (sokuon) are studied using the large vocabulary speech database.

An auditory pitch detection model based on narrowed autocoincidence histogram has been studied.

Technical Publications: [Cheveigne-89-10a, 10c], [Related technical reports are also available on request]

2.8. Speech Database

A speech database with phoneme labels is essential and necessary not only for speech recognition research but also for speech synthesis research.

2.8.1. Large-Scale Speech Database with Phonetic Transcriptions [Takeda, Kuwabara, M. Abe, Sagisaka, Ohkura, Umeda]

A large-scale Japanese speech database has been constructed at ATR with the help of several other laboratories. For multiple transcriptions, three types of categories are considered: phonetic labels, acoustic events and allophonic variations. To date, about 8,500 words uttered by twenty professional announcers and four ordinary speakers have been collected and transcribed. The speech database has effectively been used to perform speech research at ATR and other laboratories. A CD-ROM of the speech database will be available in March 1990. This speech database is now being accepted as the standard.

A continuous utterance speech database with phonetic transcription, in which phoneme balance is considered from the viewpoint of entropy of CV and VC occurrences, has been constructed. Prosody labels have been tried to put on the continuous utterance speech database.

A speaker independence speech database has just been started, and a noisy speech database including additive noises and reverberation environment noises.will also be started shortly. For a better speech synthesis by rule system, a speech synthesis database and pitch database has been developed. An English speech database is also being developed uthrough collaboration with Edinburgh University and Carnegie-Mellon University.

Technical Publications: [Kurematsu-89-05, M.Abe-89-10], [Related ATR technical reports are available on request]

2.8.2. Speech Database Retrieval System [Takeda]

To easily access the speech database, a database management system was developed. The system was implemented using the relational database *UNIFY*. The easy access language, *EAL*, was implemented using the host language interface on *UNIFY*. *EAL* is now widely used at ATR.

Technical Publications: [A technical report on EAL is available on request]

2.8.3. Speech Workbench [Tanaka, Kitagaito, Cheveigne, Poser, M.Nakamura]

A speech workbench for speech researchers has been developed on a micro Vax station with an AD/DA device based on an x11-window. The main features are high quality spectrograms, easy extension of functions and easy access to the speech database. A neural network workbench is also developed.

Technical Publications: [Cheveigne-89-10b, M.Nakamura-89-10], [Related ATR technical reports are available on request]

3. Research Staff

The research staff is mainly composed of members from the research institute and laboratories which support ATR, and visiting scientists. They are listed below:

Dr. Kiyohiro Shikano,	*Department Head*	*(NTT,*	*1986.6 −)*
Dr. Hisao Kuwabara,	*Supervisor*	*(NHK,*	*1986.10− 1989.6)*
Tetsuo Umeda,	*Supervisor*	*(NHK,*	*1989.7−)*
Dr. Yoshinori Sagisaka,	*Senior Researcher*	*(NTT,*	*1986.4 −)*
Dr. Takeshi Kawabata,	*Senior Researcher*	*(NTT,*	*1986.9 −)*
Dr. Hidefumi Sawai,	*Senior Researcher*	*(Ricoh,*	*1988.4 −)*
Dr. Alex Waibel,	*Invited Researcher*	*(Carnegie-Mellon,*	*1987.5 − 1988.8*
			1989.6 − 1989.8)
Dr. William Poser,	*Invited Researcher*	*(Stanford Univ.,*	*1988.9 − 1989.2)*
Shin'ichi Tamura,	*Researcher*	*(Sony,*	*1986.9 −)*
Kaichiro Hatazaki,	*Researcher*	*(NEC,*	*1986.12− 1989.3)*
Masanori Miyatake,	*Researcher*	*(Sanyo,*	*1986.9 − 1989.9)*
Kazumi Ohkura,	*Researcher*	*(Sanyo,*	*1989.9 −)*
Masami Nakamura,	*Researcher*	*(Sumitomo Metal,*	*1987.9 −)*
Yasuhiro Komori,	*Researcher*	*(Canon,*	*1988.9 −)*
Katsuo Abe,	*Researcher*	*(Toyocom,*	*1987.3 −)*
Satoshi Nakamura,	*Researcher*	*(Sharp,*	*1986.9 −)*
Masanobu Abe,	*Researcher*	*(NTT,*	*1987.4 −)*
Kazuya Takeda,	*Researcher*	*(KDD,*	*1986.8 −)*
Katsuteru Maruyama,	*Researcher*	*(Nitsuko,*	*1987.3 −)*
Toshiyuki Hanazawa,	*Researcher*	*(Mitsubishi Elec.,*	*1987.3 −)*
Hiroaki Hattori,	*Researcher*	*(NEC,*	*1989.5 −)*

Dr. Alain de Cheveigne,	*Visiting Researcher*	*(CNRS-Univ. Paris 7,*	*1989.2 −)*
Takaharu Tanaka,	*Engineer*	*(DEC Japan,*	*1988.9 − 1989.8)*
Kouji Kitagaito,	*Engineer*	*(DEC Japan,*	*1989.9 −)*
Jean-Claude Dang,	*Intern student*	*(ENST, France,*	*1989.4 −)*
Yasuhiro Minami,	*Intern student*	*(Keio Univ.,*	*1989.4 −)*

We hope to have more visiting scientists and intern students from abroad, because our project definitely needs international research collaboration.

4. Research Facilities

The Speech Processing Department facilities include a number of computer systems which are connected through Ethernet. The computer systems include a Vax 8600, a Vax 8700, a Vax 8800 and an Alliant 9800 (eight-CPU array processor). Every researcher also has a micro Vax station or a C Vax with an AD/DA device. Another five Vax station 3100s are used mainly for neural network and HMM research. Four Masscomps (MC5600) and one micro Vax II which are equipped with 16 bit AD conversion devices are used to collect speech data. Four Symbolics (S3670s) with *ART* are used for implementing a phoneme segmentation expert system and a knowledge-base speech synthesis-by-rule system.

Technical Publication List

at the Speech Processing Department, ATR Interpreting Telephony Research Laboratories.

January, 1989 through October, 1989

(Abbreviations in the technical publication list)

ASA : The Acoustical Society of America

ASJ : The Acoustical Society of Japan

IEICE : The Institute of Electronics, Information and Communication Engineers (Japan)

ICASSP : IEEE International Conference on Acoustics, Speech, and Signal Processing

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Cheveigné -89-10a	**Narrowed Autocoincidence of Nerve Spike Patterns and Pure Tone Pitch**	A. Cheveigné H. Kuwabara	ASJ Fall meeting, 1-2-4,Toyama, (1989-10)	1
Cheveigné -89-10b	**The MapSignal Remote Speech Editor**	A. Cheveigné	ASJ Fall meeting, 3-4-15, Toyama, (1989-10)	3
Cheveigné -89-10c	**Pitch, and the Narrowed Autocoincidence Histogram**	A. Cheveigné	1st International Conference on Music Perception and Cognition (ICMPC), Kyoto,(1989-10)	5
Dang -89-10	**Deterministic Boltzmann Machines for Phoneme Recognition**	J-C. Dang H. Sawai	ASJ Fall meeting, 1-1-23, Toyama, (1989-10)	9
Endo -89-08	ニューラルネットワークによる予測モデルを用いた音韻認識 (Phoneme Recognition Experiment by Neural Prediction Models)	T.Endo (Waseda Univ.) S.Tamura M.Nakamura	ATR Technical Report TR-I-0107, (1989-08)	—
Haffner -89-03a	**Fast Back-Propagation Learning Methods for Large Phonemic Neural Networks**	P. Haffner H. Sawai A. Waibel K. Shikano	ASJ Spring meeting, 1-6-14, Waseda, (1989-03)	11
Haffner -89-03b	**Fast Back-Propagation Learning Methods for Neural Networks in Speech Recognition**	P. Haffner H. Sawai A. Waibel K. Shikano	IEICE Spring meeting, SA-1-1, Kinki Univ., (1989-03)	13
Haffner -89-09	**Fast Back-Propagation Learning Methods for Large Phonemic Neural Networks**	P. Haffner H. Sawai A. Waibel K. Shikano	European Conference on Speech Communica-tion and Technology, pp553-556, Paris, (1989-09)	15
Hanazawa -89-03	HMM音韻モデル文節認識による評価 (Evaluation of HMM Phone Units through Japanese Phrase Recognition)	T. Hanazawa K. Kita T. Kawabata K. Shikano	ASJ Spring meeting, 3-6-6, Waseda, (1989-03)	19
Hanazawa -89-10a	ベクトル量子化話者適応アルゴリズムのHMM文節認識による評価 (Evaluation of VQ Based Speaker Adapta-tion Algorithm through HMM Japanese Phrase Recognition)	T. Hanazawa S. Nakamura T. Kawabata K. Shikano	ASJ Fall meeting, 2-P-(18), Toyama, (1989-10)	21
Hanazawa -89-10b	Hidden Markovモデルによる日本語有声破裂音の認識 (Recognition of Japanese Voiced Stops Using Hidden Markov Models)	T.Hanazawa T.Kawabata K.Shikano	Journal of ASJ, Vol.10, No.10, pp776-785 (1989-10)	23

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Hatazaki -89-03a	スペクトログラム.リーディング知識による音韻セグメンテーションの評価 (Evaluation of Phoneme Segmentation Using Spectrgram Reading Knowledge)	K. Hatazaki Y. Komori	ASJ Spring meeting, 2-P-(2), Waseda, (1989-03)	33
Hatazaki -89-03b	スペクトログラム.リーディング知識を用いた音韻セグメンテーション・エキスパートシステム (Phoneme Segmentation Expert System Using Spectrogram Reading Knowledge)	K. Hatazaki Y. Komori	ATR Technical Report TR-I-0072, (1989-03)	—
Hatazaki -89-03c	スペクトログラム・リーディング知識に基づく音韻セグメンテーション知識 (Phoneme Segmentation Knowledge Based on Spectrogram Reading)	K. Hatazaki Y. Komori	ATR Technical Report TR-I-0073, (1989-03)	—
Hatazaki -89-05	**Phoneme Segmentation Using Spectrogram Reading Knowledge**	K. Hatazaki Y.Komori T. Kawabata K. Shikano	ICASSP'89, S.8.2, pp393-396 Glasgow, (1989-05)	35
Hirata -89-03	HMM-LR法を用いた文節認識における継続時間長制御パラメータ変換法の検討 (Study on Duration Control for HMM-LR Continuous Speech Recognition)	Y.Hirata (Toyohashi Tech.) T.Kawabata T.Hanazawa	ATR Technical Report TR-I-0076, (1989-03)	—
K.Abe -89-03	波形重ね合わせ法による合成音の品質について (Quality Evaluation for Synthesized Speech Using Wave Overlap Adding)	K. Abe Y. Sagisaka H. Kuwabara	ASJ Spring meeting, 1-7-16, Waseda, (1989-03)	39
K.Abe -89-10	音声合成素片の接続方法の検討 (A Study on Speech Synthesis Units Concatenation)	K. Abe Y. Sagisaka K. Takeda	ASJ Fall meeting, 3-P-5, Toyama, (1989-10)	41
Kawabata -89-03	音韻パープレキシティの提案 (Task Entropy and Phone Perplexity)	T. Kawabata K. Shikano K. Kita	ASJ Spring meeting, 3-6-12, Waseda, (1989-03)	43
Kawabata -89-05	**Island-Driven Continuous Speech Recognizer Using Phoneme-Based HMM Word Spotting**	T.Kawabata K. Shikano	ICASSP'89, S9.7, pp461-464, Glasgow, (1989-05)	45
Kawabata -89-10	構成的ニューラルネットによる音声認識 (Constructive Neural Network for Speech Recognition)	T. Kawabata	ASJ Fall meeting, 1-1-26, Toyama, (1989-10)	49
Kita -89-03	HMM-LR 連続音声認識システムにおける計算量削減のための一検討 (Computing Amount Reduction in HMM-LR Continuous Speech Recognition System)	K. Kita T. Kawabata T. Morimoto	ASJ Spring meeting, 3-6-4, Waseda, (1989-03)	51

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Kita -89-05	**HMM Continuous Speech Recognition Using Predictive LR Parsing**	K. Kita T. Kawabata H. Saito	ICASSP'89, S13.3, pp703-706, Glasgow, (1989-05)	53
Kita -89-06	HMM音韻認識と拡張LR構文解析法を用いた連続音韻認識 (HMM Continuous Speech Recognition Using Generalized LR Parsing)	K. Kita T. Kawabata H. Saito	ATR Technical Report TR-I-0082, (1989-06)	—
Kita -89-08	**Parsing Continuous Speech by HMM-LR Method**	K. Kita T. Kawabata H. Saito	International Workshop on Parsing Technologies, pp126-131 Pittsburgh, (1989-08)	57
Kita -89-10	SL-TRANSにおける文節音声識 (Speech Recognition Method in SL-TRANS)	K. Kita T. Sakano J. Hosaka T. Kawabata	Information Processing Society of Japan, Fall meeting, Kyusyu Institute of Technology, (1989-10)	63
Komori -89-03	スペクトログラム.リーディング知識に基づく音韻認識エキスパートシステムの構築 (Phoneme Recognition Expert System Using Spectrogram Reading Knowledge)	Y. Komori K. Hatazaki T. Tanaka T. Kawabata	ASJ Spring meeting, 3-6-11, Waseda, (1989-03)	65
Komori -89-06	スペクトログラム.リーデイング知識とニューラル.ネットワークを用いた音韻認識エキスパートシステム **(Phoneme Recognition Expert Spectrogram Reading Knowledge and Neural Networks)**	Y. Komori K. Hatazaki T. Tanaka T. Kawabata K. Shikano	IEICE Technical Report, SP89-33 , Iwate Univ, (1989-06)	67
Komori -89-09	**Phoneme Recognition Expert System Using Spectrogram Reading Knowledge and Neural Networks**	Y. Komori K. Hatazaki T. Tanaka T. Kawabata K.Shikano	European Conference on Speech Communication and Technology, pp549-552, Paris, (1989-09)	75
Komori -89-10	スペクトログラム.リーディング知識に基づく音韻認識エキスパートシステムにおける音韻識別ニューラルネットワークの融合法の検討 (Combining Phoneme Identification Neural Networks into an Expert System Using Spectrogram Reading Knowledge)	Y. Komori K. Hatazaki T. Tanaka T. Kawabata K. Shikano	ASJ Fall meeting, 3-1-14, Toyama, (1989-10)	79
Kurematsu -89-09	**ATR Japanese Speech Database as a Tool of Speech Recognition and Synthesis**	A. Kurematsu K. Takeda H. Kuwabara K. Shikano	ESCA-Workshop on Speech I/O Assessment, 2.3.1, Netherlands, (1989-09)	81

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Kuwabara -89-05	**Construction of a Large-Scale Japanese Speech Database and its Management System**	H. Kuwabara K. Takeda Y. Sagisaka S. Morikawa(TIS) T. Watanabe(NEC)	ICASSP'89 S10b.12, pp560-563, Glasgow, (1989-05)	85
Kuwabara -89-07a	日本語促音の特徴 (Characteristics of a Japanese Phoneme "Sokuon")	H. Kuwabara N. Yoshida (Kyoto Univ.)	ATR Technical Report TR-I-0087, (1989-07)	—
Kuwabara -89-07b	日本語におけるアクセントの機能と音韻句の問題 (Accentuation of Japanese Language and Problems of Phonological Phrase)	H. Kuwabara Y. Iwai (Kyoto Univ.)	ATR Technical Report TR-I-0088, (1989-07)	—
Kuwabara -89-07c	研究用ATR音声データベースの作成 (Construction of ATR Japanese Speech Database as a Research Tool)	H. Kuwabara M. Abe Y. Sagisaka K. Takeda	ATR Technical Report TR-I-0086, (1989-07)	—
M.Abe -89-03	入力スペクトル情報の利用による声質変換法の高度化 (Improvement of Voice Conversion by Usage of Input Speech Spectrum)	M. Abe S. Nakamura K. Shikano H. Kuwabara	ASJ Spring meeting, 1-7-23, Waseda, (1989-03)	89
M.Abe -89-04a	連続音声の基本周波数データベース (Pitch Frequency Database on Continuous Speech)	M. Abe H. Kuwabara	ATR Technical Report TR-I-0078, (1989-04)	—
M.Abe -89-04b	連続音声データベースにおける言語.韻律情報 (Integrating Linguistic and Prosodic Information in a Continuous Speech Database)	M. Abe Y. Sagisaka H. Kuwabara	ATR Technical Report TR-I-0079, (1989-04)	—
M.Abe -89-05	**A New Speech Modification Method by Signal Reconstruction**	M. Abe S. Tamura H. Kuwabara	ICASSP'89, S11.9, pp592-595 Glasgow, (1989-05)	91
M.Abe -89-08	FFTスペクトルからの信号再生法による音声変換手法 (A Speech Modification Method by Sinal Reconstruction Using Short-time Fourier Transform)	M. Abe S. Tamura H. Kuwabara	Journal of IEICE (D-II), Vol.J72-D-II, No.8, pp1180-1186, (1989-08)	95
M.Abe -89-10	言語.韻律情報を持つ連続音声の基本周波数データベース (The Integration of Linguistic, Prosodic Information and Fundamental Frequency in a Continuous Speech Database)	M. Abe Y. Sagisaka H. Kuwabara	ASJ Fall meeting, 2-3-22, Toyama, (1989-10)	102
M.Nakamura -89-05	**English Word Category Prediction Based on Neural Networks**	M. Nakamura K. Shikano	ICASSP'89, S13.10, pp731-734, Glasgow, (1989-05)	104

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
M.Nakamura -89-09	ニューラルネット開発用ワークベンチシステムネットワークエディタおよびモニタ機能について (Neural Net Workbench System)	M. Nakamura S. Tamura M. Miyatake H. Sawai	ATR Technical Report TR-I-0113, (1989-09)	—
M.Nakamura -89-10	ニューラルネット開発用ワークベンチシステムネットワークエディタおよびモニタ機能について (Neural Net Workbench System)	M. Nakamura S. Tamura M. Miyatake H. Sawai	ASJ Fall meeting, 2-P-(26), Toyama, (1989-10)	108
Maruyama -89-01	HMM音韻連結学習を用いた英単語音声の認識 **(English Word Recognition Using HMM Phone Concatenated Training)**	K. Maruyama T. Hanazawa T. Kawabata K. Shikano	IEICE Technical Report, Osaka Univ., SP88-119, (1989-01)	110
Maruyama -89-03	HMM音韻連結学習を用いた英単語音声の認識 (English Word Recognition Using HMM Phone Concatenated Training)	K. Maruyama T. Hanazawa T. Kawabata K. Shikano	ASJ Spring meeting, 1-6-22, Waseda, (1989-03)	117
Maruyama -89-09	高精度HMMを用いた英単語認識 (English Word Recognition Using Multiple Codebooks)	K. Maruyama T. Hanazawa T. Kawabata K. Shikano	ATR Technical Report TR-I-0108, (1989-09)	—
Maruyama -89-10	HMM音韻連結学習とNETgramを用いた英単語音声の認識 (English Word Recognition Using HMM Phone Concatenated Training and NETgram)	K. Maruyama M. Nakamura T. Kawabata K. Shikano	ASJ Fall meeting, 2-P-(8), Toyama, (1989-10)	119
Minami -89-07	TDNN音韻スポッティングと拡張LRパーザを用いた文節音声認識 (Continuous Speech Recogniton Using TDNN Phoneme Spotting and Generalized LR Parser)	Y. Minami M. Miyatake H. Sawai K. Shikano	ATR Technical Report TR-I-0085, (1989-07)	—
Minami -89-10	TDNN音韻スポッティングと拡張LRパーザを用いた分節音声認識 (Continuous Speech Recognition Using TDNN Phoneme Spotting and Generalilzed LR Parser)	Y. Minami M. Miyatake H. Sawai K. Shikano	ASJ Fall meeting, 3-1-11, Toyama, (1989-10)	121
Miyatake -89-03a	時間遅れニューラルネットワークを用いた音韻スポッティング法 (Phoneme Spotting Methods Using Time-Delay Neural Networks)	M. Miyatake H. Sawai K. Shikano	IEICE Spring meeting, SA-1-4, Kinki Univ., (1989-03)	123
Miyatake -89-03b	全音韻を統合した時間遅れ神経回路網(TDNN)による音韻スポッティング (Spotting Phonemes Using Integrated Time-Delay Neural Networks (TDNN))	M. Miyatake H. Sawai K. Shikano	ASJ Spring meeting, 2-P-(24), Waseda, (1989-03)	125

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Miyatake -89-06	連続音声中のスポッティングのためのTDNN構成法 **(How to Apply TDNN to Phoneme Spotting in Continuous Speech)**	M. Miyatake H. Sawai K. Shikano	IEICE Technical Report, SP89-32, Iwate Univ., (1989-06)	127
Miyatake -89-08a	時間遅れ神経回路網(TDNN)による音韻スポッティングのための効果的学習法 (Effective Training Methods for Spotting Japanese Phonemes Using Time-Delay Neural Networks)	M. Miyatake H. Sawai K. Shikano	ATR Technical Report TR-I-0103, (1989-08)	—
Miyatake -89-08b	音声研究用ワークベンチの処理の自動化 (An Automatic Hardcopy Tool Based on Speech Workbench)	M. Miyatake Y. Sagisaka	ATR Technical Report TR-I-0106.,(1989-08)	—
Miyatake -89-08c	会話文音声生成のための音声合成、およびニューラルネットワークの連続音声認識への適用 (Prosody Control for Conversational Speech Synthesis, and Neural Networks for Continuous Speech Recognition)	M. Miyatake	ATR Technical Report TR-I-0112.,(1989-08)	—
Miyatake -89-10	時間遅れ神経回路網(TDNN)による音韻スポッティングの改良 (Improvement on Spotting Phonemes Using Time-Delay Neural Networks (TDNN))	M. Miyatake H. Sawai K. Shikano	ASJ Fall meeting, 1-1-25, Toyama, (1989-10)	133
Poser -89-01a	**On Phrasal Downtrends in F0 in Japanese**	W. Poser	Kinki Society for Phonetics Meeting, (1989-01)	135
Poser -89-01b	**Modified MITalk**	W. Poser M. Abe	ATR Technical Report TR-I-0066.,(1989-01)	—
Poser -89-03a	**Implementation of an F0 Model for Japanese Incorporating Downstep**	W. Poser Y. Sagisaka	ASJ Spring meeting, 1-7-5, Waseda, (1989-03)	150
Poser -89-03b	**Tools for Fundamental Frequency Modelling**	W. Poser	ATR Technical Report TR-I-0069, (1989-03)	—
Poser -89-03c	**Modelling Phrasal Level F0 Phenomena in Japanese**	W. Poser Y. Sagisaka	IEICE Technical Report,SP88-160, Tokyo, (1989-03)	152
S.Nakamura -89-02	ファジィベクトル量子化によるスペクトログラムの正規化 (Spectrogram Normalization Using Fuzzy Vector Quantization)	S.Nakamura K.Shikano	Journal of ASJ, Vol.45, No.2, pp107-114, (1989-02)	158

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
S.Nakamura -89-03	ベクトル量子化話者適応化アルゴリズムのHMM音韻認識による評価 (Phoneme Recognition Evaluation of HMM Speaker Adaptation Using Vector Quantization)	S. Nakamura T. Hanazawa K. Shikano	ASJ Spring meeting, 1-6-23, Waseda, (1989-03)	166
S.Nakamura -89-05	**Speaker Adaptation Applied to HMM and Neural Networks**	S. Nakamura K. Shikano	ICASSP'89, S3.3, pp89-92, Glasgow, (1989-05)	168
S.Nakamura -89-06	時間遅れ神経回路網(TDNN)における入力パラメータの評価と話者適応化 **(Comparison of Input Parameters for Time-Delay Neural Network(TDNN) and Implementation of VQ-Based Speaker Adaptation)**	S. Nakamura K. Shikano	IEICE Technical Report, SP89-18, Iwate Univ., (1989-06)	172
S.Nakamura -89-08a	セパレートベクトル量子化に基づく話者適応化 (Speaker Adaptation through Separate Vector Quantization)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0095, (1989-08)	—
S.Nakamura -89-08b	ファジイベクトル量子化に基づく話者適応化 (Speaker Adaptation through Fuzzy Vector Quantization)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0096, (1989-08)	—
S.Nakamura -89-08c	ベクトル量子化話者適応のHMM音韻認識への適応 (VQ-Based Speaker Adaptation Applied to HMM Phoneme Recognition)	S. Nakamura T. Hanazawa K. Shikano	ATR Technical Report TR-I-0097, (1989-08)	—
S.Nakamura -89-08d	ベクトル量子化話者適応の時間遅れ神経回路網による音韻認識への適用 (VQ-Based Speaker Adaptation Applied to Time-Delay Neural Network Phoneme Recognition)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0098, (1989-08)	—
S.Nakamura -89-08e	話者重畳型HMMを用いた調音様式の話者適応化 (Speaker Adaptation of Articulatory Variation by Supplemented HMM)	S. Nakamura H. Hattori K. Shikano	ATR Technical Report TR-I-0099, (1989-08)	—
S.Nakamura -89-08f	ベクトル量子化話者適応化の研究 (A Study of VQ-based Speaker Adaptation)	S. Nakamura	ATR Technical Report TR-I-0100, (1989-08)	—
S.Nakamura -89-08g	話者適応化における写像方法の比較 (A Comparative Study of Spectral Mapping Methods on Speaker Adaptation)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0101, (1989-08)	—
S.Nakamura -89-10a	複数話者HMM学習を用いた話者適応化の音韻認識による評価 (A Study for Supplementation of Speaker Articulatory Variation on HMM Speaker Adaptation)	S. Nakamura H. Hattori K. Shikano	ASJ Fall meeting, 2-P-(15), Toyama, (1989-10)	180

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
S.Nakamura -89-10b	教師つき話者適応化における写像方法の比較 (A Comparative Study of Spectral Mapping Methods on Supervised Speaker Adaptation)	S. Nakamura K. Shikano	ASJ Fall meeting, 2-P-(14), Toyama, (1989-10)	182
Sagisaka -89-03	情報量尺度を用いた音声単位セットの構成法 (On the Design of a Speech Synthesis Unit Set Using Entropy Measure)	Y. Sagisaka	ASJ Spring meeting, 1-7-20, Waseda, (1989-03)	184
Sagisaka -89-05	**The Integration of Rules and Data in a Speech Synthesis System**	Y. Sagisaka	Speech Tech'89 New York, (1989-05)	186
Sagisaka -89-10	統語構造に基づくF0パタン概形の制御 (On the Control of Global F0 Pattern Using Simple Syntactic Information)	Y. Sagisaka	ASJ Fall meeting, 3-P-13, Toyama, (1989-10)	189
Sawai -89-03	**Spotting Phonemes by Hierarchical Construction of Time-Delay Neural Networks**	H. Sawai M. Miyatake K. Shikano	ASJ Spring meeting, 2-P-(25), Waseda, (1989-03)	191
Sawai -89-05	**Spotting Japanese CV-Syllables and Phonemes Using Time-Delay Nerual Networks**	H. Sawai A. Waibel M. Miyatake K. Shikano	ICASSP'89 S.1.7, pp25-28, Glasgow, (1989-05)	193
Sawai -89-06	**Parallelism, Hierarchy, Scaling in Time-Delay Neural Networks for Spotting Japanese Phonemes/CV-Syllables**	H. Sawai A. Waibel P. Haffner M. Miyatake K. Shikano	IEEE International Joint Conference on Neural Networks, (IEEE/INNS) Washington, D.C., (1989-06)	197
Sawai -89-07a	時間遅れ神経回路網を用いた音韻/音節スッポッティング (Spotting Phonemes and Syllables using Time-Delay Neural Networks)	H. Sawai M. Miyatake A. Waibel K. Shikano	ATR Technical Report TR-I-0089, (1989-07)	—
Sawai -89-07b	**Parallelism, Hierarchy, Scaling in Time-Delay Neural Networks for Spotting Phonemes and CV-Syllables**	H. Sawai A. Waibel P. Haffner M. Miyatake K. Shikano	ATR Technical Report TR-I-0090, (1989-07)	—
Sawai -89-08	連続音声認識のための時間遅れ神経回路網を用いた音韻/音節スポッティング (Spotting Phonemes and Syllables for Continuous Speech Recognition Using Time-Delay Neural Networks)	H. Sawai M. Miyatake A. Waibel K. Shikano	Journal of IEICE,(D II), invited paper, Vol.J72-D-II, No.8, pp1151-1158, (1989-08)	205

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Seiyama -89-10	複素ケプストラムによる周期成分波形を零位相化することによる基本周波数抽出法 (A Pitch Extraction Method Using Phase Equalization of Impulse Train by Complex Cepstrum)	N.Seiyama (NHK) T.Umeda	ASJ Fall meeting, 2-3-14, Toyama, (1989-10)	213
Shikano -89-01	**Research Activities of the Speech Processing Department (October 1987 through December 1988)**	K. Shikano	ATR Technical Report TR-I-0065, (1989-01)	—
Shikano -89-05	ニューラルネットワークの音声情報処理への応用 (Neural Networks Applied to Speech Processing)	K. Shikano	Seminar of Information Processing Society in Kansai, pp67-79, (1989-05)	215
Shikano -89-09	音韻認識エキスパートシステム (Phoneme Recognition Expert System)	K. Shikano K. Hatazaki Y. Komori	IEICE Fall meeting, PD-1-4, Kanagawa Univ., (1989-09)	228
Takagi -89-10	声質変換による、感情表現制御のための物理的要因の解析 (Analysis of Acoustic Features for Control of Emotional Expressions of Speech by Voice-Quality Conversion)	T. Takagi (NHK) H. Kuwabara (NHK) T.Umeda	ASJ Fall meeting, 3-3-4, Toyama, (1989-10)	230
Takeda -89-05	**An Approach for Normalizing Coarticulatory Variations through Expansion of Acoustic Properties into Phonetic Features**	K. Takeda	ATR Technical Report TR-I-0081, (1989-05)	—
Takeda -89-09	**Adaptive Manipulation of Non-Uniform Synthesis Units Using Multi-level Unit Transcription**	K. Takeda K. Abe Y. Sagisaka H. Kuwabara	European Conference on Speech Communication and Technology, pp195-198, Paris, (1989-09)	232
Takeda -89-10	合成単位の選択における音韻的知識と韻律的属性の処理の統合 (Unit Selection Measurement Combining Phonetic and Prosodic Criteria for Synthesis-by-Rule)	K. Takeda Y. Sagisaka H. Kuwabara	ASJ Fall meeting, 3-P-16, Toyama, (1989-10)	236
Tamura -89-01	**An Analysis of a Noise Reduction Neural Network which Takes Waveforms as Input and Output**	S. Tamura	IEICE Technical Report, SP88-124, (1989-01)	238
Tamura -89-05	**An Analysis of a Noise Reduction Neural Network**	S. Tamura	ICASSP'89, A.1a.6, pp2001-2004, Glasgow, (1989-05)	246
Tamura -89-06	**On Interpretations of a Feed-Forward Neural Network**	S. Tamura	IEEE International Joint Conference on Neural Networks ,(IEEE/INNS), II-584, Washington, D.C., (1989-06)	250

Technical Publication List at the Speech Processing Department (January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Tamura -89-09	ニューラルネットを用いた音声処理 (Neural Networks Applied to Speech Processing)	S. Tamura H. Sawai M. Nakamura K. Shikano	The Journal of the Institute of Television Engineers of Japan, Vol.43, No.9, pp935-943, (1989-09)	258
Tanaka -89-08a	X Window System, Version II ポケット.ガイドーXlib編ー (X Window System, Version II, Pocket Guide - Xlib -)	T. Tanaka	ATR Technical Report TR-I-0091, (1989-08)	—
Tanaka -89-08b	X Window System, Version II ポケットガイドーTool Kit編ー (X Window System, Version II, Pocket Guide - Tool Kit -)	T. Tanaka	ATR Technical Report TR-I-0092, (1989-08)	—
Waibel -89-03	**Phoneme Recognition Using Time-Delay Neural Networks**	A. Waibel T.Hanazawa G.Hinton (Tronto Univ.) K.Shikano K.Lang (CMU)	IEEE Tr.ASSP, Vol.37, No.3, pp328-339, (1989-03)	271
Waibel -89-05	**Consonant Recognition by Modular Construction of Large Phonemic Time-Delay Neural Networks**	A. Waibel H. Sawai K. Shikano	ICASSP'89, S.3.9, pp112-115, Glasgow, (1989-05)	267