

Speech Processing Technologies

Prospects for Speech Interface Technologies

FURUI Sadaaki

Speech Processing Technologies Becoming Common in Daily Life, and Toshiba's Approach

AKAMINE Masami

Speech interfaces have recently become increasingly widespread for interacting with digital devices such as smartphones instead of touch keyboards.

Since the 1980s, Toshiba has been developing various core technologies supporting speech interfaces, such as automatic speech recognition, speech synthesis, and so on. These technologies have been applied to a variety of products including speech middleware for in-car navigation systems, dictation software, content-creation services on websites, and machine translation systems. Aiming at the realization of the so-called cognitive assistant, we have been continuously engaged in the development of not only speech technologies but also technologies related to multimodal interfaces and various new products and services.

Large-Vocabulary Speech Recognition Technologies for Achievement of Simultaneous Translation and Speech Dialog Systems

MASUKO Takashi / ASHIKAWA Masayuki

In order to achieve the practical use of voice translation and speech dialog systems, large-vocabulary speech recognition that recognizes utterances of various types is required. However, it is difficult for a small number of developers to collect the new words and colloquial expressions that continuously appear in the language and to add them to a system's vocabulary.

Moreover, it is necessary to improve phoneme discrimination performance in order to discriminate between the increasing number of similarly pronounced words that emerge with the expansion of vocabulary size.

To overcome these problems, Toshiba has established a word collection method using crowdsourcing and developed a new acoustic feature to improve phoneme discrimination ability. These technologies realize large-vocabulary speech recognition through the collection of a number of words in a short period of time and improved speech recognition accuracy.

Text-to-Speech Technologies Realizing Various Voices and Expressive Reading

MORITA Masahiro / TAMURA Masatsune / FUME Kosei

As text-to-speech (TTS) technologies are now widely used for e-book reading and entertainment applications, improvement of their ability to provide various types of voices, speaking styles, and emotions has become a focus of attention.

In response to this need, Toshiba has developed the following advanced TTS technologies: (1) a custom voice production technology that can build a wide variety of voices closely resembling the voices of specific people at low cost and within a short time; (2) an expressive reading technology that can automatically select emotions from respective dialogues in such works as novels; (3) a prosodic authoring technology that can efficiently create speech contents with the intended intonation; and (4) a digital watermarking technology that prevents the misuse of TTS, such as for identity theft.

Spoken Dialogue Technology to Understand Problems and Offer Solutions

NAGAE Hisayoshi / YAMASAKI Tomohiro / ICHIMURA Yumi

Spoken dialogue systems such as personal assistant applications for smartphones have appeared in recent years. In order to hold a meaningful conversation with a conventional personal assistant application, however, it is necessary to input sentences containing explicit commands. Attention has therefore been increasingly focused on a spoken dialogue system to which users can speak freely, without the need for predetermined commands.

Toshiba has developed a spoken dialogue technology that can assist in resolving users' problems through more spontaneous human-machine dialogues. This technology makes it possible to provide adequate solutions to users through the estimation of intended meaning based on background knowledge collected from large amounts of data, including words and patterns in sentences, even when a user utters an ambiguous expression rather than giving a clear instruction to the system.

Simultaneous Interpretation Technology Supporting Conversations in Foreign Languages for Face-to-Face Services

KAMATANI Satoshi / SAKAMOTO Akiko / SUMITA Kazuo

With the increasing opportunities for conversation in foreign languages, demand has been growing for a simultaneous machine interpretation technology that can be used in many different situations.

Toshiba has developed a simultaneous interpretation system for continuous speech conversation taking place in various face-to-face services at stores, reception desks, counters at public offices, and so on. This system, is capable of both Japanese/English and Japanese/Chinese interpretation, supports smoother communication between speakers of different languages by processing their continuous spontaneous speech and incrementally outputting the interpretation results. As a result, a user can immediately understand what a conversational partner is saying. We have conducted field experiments and confirmed that a solved task ratio of approximately 90% is achieved for various tasks including buying souvenirs and asking for directions regarding a bus route.

High-Quality Voice Capture Technologies and Application to Tablet

ISAKA Takehiko / SUDO Takashi / AMADA Tadashi

Demand has been increasing for voice input applications including video chat systems and speech recognition systems. To improve the usability of these applications, it is essential to capture voices as clearly as possible.

In order to minimize factors that degrade quality in voice input applications, Toshiba has developed the following high-quality voice capture technologies: (1) an echo canceller to suppress sounds from a speaker being picked up by a microphone, (2) beamforming to suppress directional noise, and (3) a noise canceller to suppress diffuse noises entering a microphone from various directions. These technologies have been implemented in the REGZA Tablet AT703/AT503 models, which feature a smooth voice capturing function.

Audio Source Separation Technology to Control Volume Balance between Voices and Background Sounds

HIROHATA Makoto / ONO Toshiyuki / NISHIYAMA Masashi

The wide dissemination of audiovisual (AV) products has provided users with easy and diversified styles of viewing and listening to video contents. However, it is not always possible to view video contents comfortably because of an imbalance in the volumes of voices and background sounds.

Toshiba has developed an audio source separation technology to extract voice and background sound source signals from audio signals. This new technology realizes a more enjoyable viewing experience by allowing users to adjust background sounds and hear voices more easily, thus providing highly realistic sensations when watching programs such as sports matches and enhancing the experience of karaoke while watching music programs.

Voice Interface for Operation of Distant Equipment

OUCHI Kazushige / KOGA Toshiyuki

In order to operate distant equipment by a speech recognition system, there are two technical challenges for the realization of practical recognition accuracy: (1) commanding the equipment to start speech recognition, and (2) reducing the influence of ambient noises.

Toshiba has developed a voice interface for the operation of distant equipment that utilizes a microphone array technology to emphasize the sound in the target direction and suppress noises from other directions. When a user activates the speech recognition system by clapping twice, the system simultaneously detects the direction of the clapping and sets the directivity angle of the microphones to that direction so as to prioritize the input of the target user's voice. We have conducted evaluation experiments using the operation of a TV set as a motif, and confirmed that users can operate a TV from 4.5 m away by means of speech recognition with a practical level of performance.

ToScribe™ Web Application to Enhance Efficiency of Audio Transcription Work

UENO Koji / ASHIKAWA Taira

Toshiba has launched ToScribe™, a new, free, cloud-based application that allows users to manually transcribe speeches more efficiently by integrating a number of speech and language processing technologies including automatic speech recognition (ASR) technology. ToScribe™ works with major Web browsers, and offers effective transcription assistance while simplifying troublesome audio player control operations by means of the following high-level speech and language processing technologies: automatic speech position estimation by manipulating the internal results of the ASR, automatic speaker estimation by clustering audio feature values, and proofreading assistance applying our text structure analysis technology.

Specular Reflection Control Technology to Increase Glossiness of Images

KOBIKI Hisashi / NONAKA Ryosuke / BABA Masahiro

The emergence of next-generation displays for televisions including 4K ultra-high-definition (Ultra HD: 3,840 x 2,160 pixels) displays and self-luminous displays using organic electroluminescent devices has enabled viewers to enjoy higher-resolution and higher-contrast contents than ever before. Under these circumstances, image processing technologies realizing images with a high-quality texture are becoming essential to take advantage of the features of these devices.

With this as a background, Toshiba has developed a specular reflection control technology to enhance the glossiness of objects in an image by using a specular reflection image separated from the input image. This technology makes it possible to optimize the required image quality for next-generation displays by incorporating our image processing technologies including a texture restoration technology.

Software Structure Diagnosis Method to Evaluate and Improve Design Maintainability

OKAMOTO Wataru

In the software development field, derivative development by changing the specifications of existing products and adding new functions to them has been increasing in recent years. Efficiency in this type of software development is dictated by maintainability, which is the capability to adapt the design of a software in this way. However, numerous modifications of a software over the long term can lead to reduced maintainability due to inadequate design changes and the resulting degradation of its structure.

To realize more efficient derivative development of software from existing software products, Toshiba has developed a software structure diagnosis method to evaluate and improve the maintainability of software design. We have defined a software metrics taking into consideration the sizes of software modules and probability of changes, and confirmed that this method can improve maintainability by identifying and displaying design problems.

1.6 Tbyte SSD for Enterprise Use Applying MLC NAND Flash Memory

KIMURA Akihiro / MORO Hiroyuki / MATSUSHITA Hiroki

The use of solid state drives (SSDs) in enterprise servers and storage systems for network computing has been expanding due to their superior performance, such as ultrahigh levels of input/output operations per second (IOPS) and IOPS per watt, compared with those of conventional hard disk drives (HDDs). However, the higher bit cost of the NAND flash memory used in SSDs compared with that of the magnetic disks used in HDDs, as well as the use of single-level cell (SLC) NAND flash memories in conventional SSDs to ensure the high data integrity required for enterprise use, are significant issues that need to be resolved in order to achieve large-capacity SSDs.

As a solution to this issue, Toshiba has developed a 2.5-inch SSD for enterprise use achieving the highest capacity in the industry of 1.6 Tbytes, applying multilevel cell (MLC) NAND flash memory. This SSD offers not only a lower bit cost but also higher performance and reliability than any previously existing models by means of a newly developed SSD controller.

Top Runner Transformers 2014 Compliant with New Criteria for Energy Consumption Efficiency

MATSUOKA Yasuhiro / KUBOTA Masaharu

The Top Runner program has been introduced in Japan to advance the energy efficiency of machinery and equipment. Under the notification of the Second Evaluation Standard of the Law Concerning the Rational Use of Energy for designated machineries including transformers, transformers shipped from April 2014 onward must be compliant with the new criteria for energy consumption efficiency, or so-called "Top Runner transformers 2014." Top Runner transformers 2014 are required to have higher energy consumption efficiency than existing Top Runner transformers.

To contribute to environmental protection and improvement of the reliability of electric power supplies, Toshiba has developed a lineup of Top Runner transformers 2014. These transformers achieve space saving due to miniaturization balanced with energy consumption efficiency, as well as enhanced aseismic performance in consideration of the damage caused by the Great East Japan Earthquake.

IBS-1000 Currency Sorter for Overseas Markets

SATO Masashi

There is an increasing need for currency processing machines in overseas markets, both to automate banknote processing due to the expansion of automated teller machines (ATMs) in industrialized countries and to accommodate the remarkable economic growth in developing countries. In response to this trend, machines with fast processing speed, a compact footprint, and system expandability are required.

Toshiba has developed the IBS-1000 currency sorter for overseas users handling large volumes of banknotes, such as commercial banks, cash-in-transit (CIT) companies, casinos, and so on. The IBS-1000 offers increased flexibility and expandability by employing a modular structure, as well as the highest level of processing speed in the industry and innovative compact on-line strapping modules. As a result, mixed-denomination notes can be automatically counted, sorted, authenticated, and strapped in one operation, thus contributing to improved operational efficiency. We have also succeeded in achieving competitive prices not only for Western markets but also expanding emerging markets by outsourcing all of the material procurement and manufacturing.