Phone Call Translator System in Real-Time

Yoko Iimura*, Yu Kojo, Masahiro Oota, and Shinya Tachimoto

NTT DOCOMO, INC.,
3-6 Hikarinooka, Yokosuka-shi, Kanagawa, 239-8536, Japan
{iimuray,kojou,ootamas,shinya.tachimoto.yw}@nttdocomo.co.jp

Abstract. Recently, various kinds of automatic translator services have become available. Most of them are provided as a client-server model. We developed a phone call translator service, which enables people speaking different languages to communicate with each other over the phone. In this paper, we present the system architecture of the service and describes how the system works.

Keywords: translation, phone, mobile application.

1 Introduction

In recent years, an automatic translation function has been provided as various software applications and web services. Most of them are offered in one-to-one setting between the user and the server and designed to indirectly support communication between different languages..

The telephone provides communication services, which have the following features:

・Natural communication experience by talking to the other person while listening to his/her voice
・Real-time communication

These conventional features of the telephone, however, make it difficult for a general telephone translation service as described above to be applied effectively to the telephone; because via the existing telephone system it is hard to achieve smooth communication between persons speaking different languages and maintain a real-time conversation at one time. To address such difficulty, we developed an automatic phone call translator system. In this paper, we describe how to implement the phone call translator system.

2 Service Overview

The phone call translator system automatically converts what the user says in Japanese into another language (English, Chinese or Korean) and vice versa.

* Corresponding author.

G. Memmi and U. Blanke (Eds.): MobiCASE 2013, LNICST 130, pp. 290–293, 2014.

The system provides translations both in screen texts and voice readouts in a synthesized voice.

Using the conventional phone lines and functions, the system allows users to hear each other's voice as they converse over the phone. This gives users a face-to-face like natural communication experience that the telephone especially can provide. In addition, the system allows users to use basic functions of the automatic translator service with anyone who can be reached by phone.

In order to enhance the convenience of the translator service, we provide a translation application (hereinafter called the translation app) operable for Android OS 2.2 or later. Figure1 shows some images of the translation app when using the service. Fusion of telecom (phone) and web (app) makes it possible for users to visually as well as aurally confirm their operation details, results of voice recognition and translation and manipulate them on the translation app. All of these capabilities offer a new style of communication.

The translation function, however, is not intended to be applied to all conversations on the phone; the function is turned on and off by the user operation. The function to start/stop the translation function is offered in two ways: one works with the translation app and the other works with the button on the phone for the user who does not use the translation app.

Fig. 1. User interface up to service usage

3 Overall Sequence

The phone call translator service is provided through a translator system that operates as a B2BUA. To activate the translator service, the user dials the special number "138" and the translation language code (2 digits) before the other party's phone number. When the special number "138" is dialed, a SIP_INVITE signal is sent to the translator system. Upon receiving the SIP_INVITE signal, the translator system calls the other party if the system determines that the user may be connected to the other party based on their service and user conditions. When the called party answers the call, the U-Plane connection is established. Then the basic function of this service is available. When using the translation app, the app makes a call instead of the user. The user just selects the language and the party to be called according to the

navigation of the app. The translator system sends a push signal to every UE to activate the translation app and connects the packet network communication from the app with the established session over the phone line. The translation app receives the results of voice recognition and translation of phone conversations via the packet communication and displays them on the screen.

4 System Configuration

We are building a network infrastructure designed to provide value added services. The infrastructure is called the Service Enabler Network (SEN), which implements AS (Application Servers) of the IMS（IP Multimedia Subsystem） infrastructure [1]. The SEN infrastructure consists of an array of different functional components and service scenarios. The policy of this infrastructure is to provide services timely by combining necessary components only through the development of service scenarios.

The phone call translator system is implemented as an AS on the SEN infrastructure stated above.

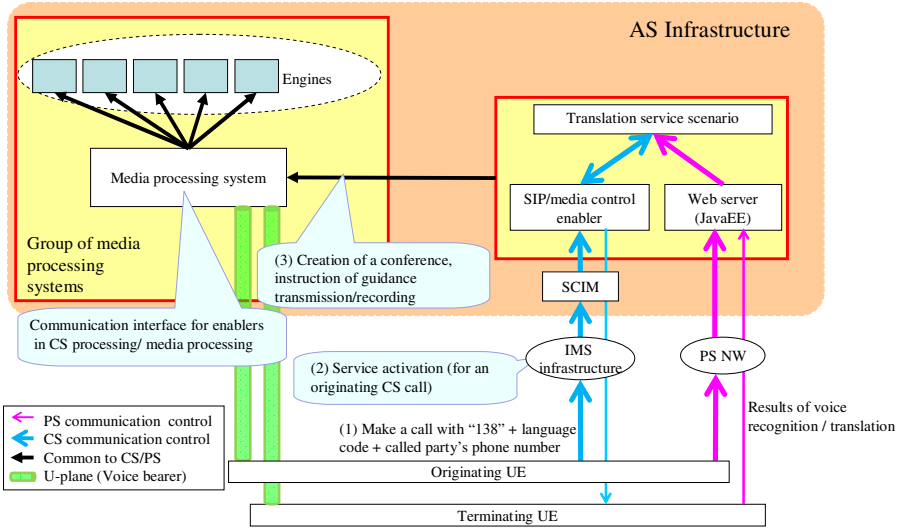

Fig. 2. System Architecture

When the user's device initiates a voice call, dialing the special number "138," the IMS infrastructure recognizes it as the service number of the phone call translator service and connects the call to the SCIM (Service Capability Interaction Manager)[2] (Figure2 (2)). When the SCIM determines that the service does not conflict with other services, it connects the voice call to the phone call translator service scenario. The scenario generates a conference in the media processing system based on the INVITE signal from the SCIM. Then it activates a conference service to be joined by the

calling and called parties as participants and draws the voice call into the media processing system. When the call starts, the scenario instructs the media processing system to send guidance to tell the start of recording and start recording the voice call (Figure 2 (3)). The media processing system records voice data and sends the guidance based on the request from the service scenario.

Each component of enablers necessary for translating speech over the phone is called an engine. For example," voice recognition engine", "Voice synthesis engine"," Translation engine" are implemented for the phone call translator service. Any gaps of engine interfaces are absorbed by the media processing system to abstract the media processing instructions from the scenario. This enables the use of the same service scenario for different engine products. It also makes it possible to select the most accurate product for each function and language and use them in combination. In addition, the system will allow us to replace an engine by more advanced one without affecting the service scenario when such need arises in the future.

5 Service Evaluation

The phone call translator service has achieved a certain level of close-to-real-time performance in its usage in Japan: for example, in case of spoken words of about 8.3 seconds in Japanese, the results of voice recognition and translation were displayed on the translation app screen about 1 second after the speech; the replay of the translated speech in a synthesized voice was started from the phone line almost at the same time.

When the service was used in areas with longer packet transmission delays, however, there were certain timing gaps between the speech of the words over the phone line and the display of the results on the screen.

6 Conclusion

In this paper, we have presented how to implement a new phone call translator system that provides new additional value for communication, enabling interactions between the phone line and the packet network.

To be able to utilize this translator system in more diverse scenes, it is essential to improve the accuracy of technical components for both voice recognition and machine translation. To this end, our future work should focus on extracting unknown words that the current engine cannot recognize from the actual translation data and list them in dictionaries as well as adopting timely words as necessary.

References

1. 3GPP TS23.228
2. 3GPP TS23.002