Language Translation as a Voice Service for Mobile Network Operators

Máté Ákos TÚNDIK, Loránd NAGY
Nokia Mobile Networks, Cloud Core
Budapest University of Technology and Economics
E-mail: {akos.tundik,lorand.1.nagy}@nokia.com

Attila HILT, Gergő BÓTA
Nokia Mobile Networks, Cloud Core
E-mail: attila.hilt@nokia.com,
gergo.bota@nokia.com

Abstract—Nokia offers a language translation solution for mobile network operators. Operators can provide real time machine translation of speech, as a useful extension of their currently available voice service. The concept provides a beneficial differentiation against Over-the-Top (OTT) content providers. The successfully demonstrated solution presented in this paper works with any device, that is capable for voice calls (even with simple handsets, like the recently launched Nokia 3310). In the service the Operator’s Mobile Switching Center Server (MSS) or IMS core is involved, which is connected to Nokia Data Center and then to the Microsoft Azure Data Center, which provides the speech translation.

Keywords—speech translation, voice call, mobile operator, core network, IMS, VoLTE

I. INTRODUCTION

Let us consider the typical situation, when people travel to countries where they do not speak the local language and for example they encounter problems with their rented car. In this case, when they do not speak the language of the visited country, it is very hard to express themselves to make their problems understandable. For example, a Russian driver has a flat tire of the rented car in England, and the customer care person of the car rental company can speak English but not Russian. How can he/she support the customer in this difficult circumstance?

Nokia provides a new language translation solution for network (NW) operators integrated into simple voice services during the calls, using an external resource, the Microsoft Azure Data Center (DC). The DC executes the translation between the selected language and the spoken language of the customer through Nokia Data Center, as it is shown in Fig.1.

It very beneficial, that this kind of operator language translation service works for all the customers, while OTT-services (e.g. Google, Facebook, Skype) can be used only between accepted contacts, and also there are some device-based limitations.

When parties do not have a common language, it’s hard to become friends/buddies in OTT communities. OTT service requires common App + connection inside the App. Users need to know and find the identity of the other party and require the same client/App in both A and B parties. Before anyone appears in the contact list, there must be already some level of interaction between the communicating parties. How can it happen without speaking a common language?

All-in-all, before communicating over the same OTT-service, people shall agree in a common language.

Compared to the above mentioned facts, Nokia Language Translation service can be used immediately. The discussion can be started without having a common language or earlier agreement on how to communicate. Furthermore, it is applicable also in use cases when the users realize only after dialling that the other party does not speak their language.

For example, foreigners call the reception of a hotel, where the hotel employee cannot help them in any of their known languages. In this case Nokia Language Translation solution would provide them the option to activate this service “on-the-fly”. With a specific DTMF, the translation service for a certain selected language could be activated. As a result, knowing the phone number is enough and there is no need to install any Apps, nor to create any accounts. It works for everybody who simply has a valid phone number and a voice capable device as shown in Fig.2.

In comparison, an OTT translation service has certain bandwidth requirements, which can be satisfied with 4G radio mostly. However, voice calls often use CS access network restricting the data access. For example, in Hungary currently there is only one operator with live VoLTE service, which is provided by Nokia Open TAPS.

When caller party ‘A’ wants to translate his/her speech for a foreign national (called party ’B’), it can be easily done by using this service. The phone can be simply put to speaker for a selected specific language pair, and the other party can hear the translation in the required language.

Calls between unknown peers are immediately possible e.g. in Business-to-Consumer (B2C) or Consumer-to-Business
Fig. 2. Mobile communication using API for translation

(C2B) use cases. Calls outside own contact list are common, and it is possible to start communication without any prior agreement/common language. It is also important to note that such solution could be well used for face-to-face conversations or even for conferencing.

II. BUSINESS CASE AND AVAILABILITY FOR END USERS

This solution is offered to Operators who intend to offer Language Translation for voice calls. This service can become an important part of business models between end-users, Operators and Nokia, providing wide-scale opportunities. Any operator voice network (mobile, fixed, VoWiFi) can apply this technology, the voice call is the only requirement for access NW. Moreover, it does not require data plan and does not consume any quota from it. Handovers are also supported (such as handover between cells or SRVCC) for Language Translation.

Therefore, the pricing remains predictable and simple, which is not obviously true in case of OTT services. OTT services use data connection, where the amount of consumed data and its price are unfortunately often unpredictable. That is especially true when the subscribers use roaming.

This service works with any voice capable device (feature phone, smart phone, soft phone, fixed phone), and Language Translation does not have an impact on battery life (while the OTT does). It is also worth mentioning, that OTT services are using the data bearer of the RAN, which does not have a QoS policy. For any access type, the voice services have bandwidth and audio quality policies thus making the voice quality for the speech recognition more appropriate.

A big advantage compared to OTT is, that the communication between operators is also possible, like combining language translation with call forwarding/transfer/conference scenarios as well. Security and the privacy are guaranteed due to the established architecture. The speech is not recorded and cannot be correlated to subscribers in any way, neither in Nokia, nor in external (e.g. Microsoft) Data Centers. Finally, this is a promising alternative for enterprises, because they often prefer operator services since those are available for all people (SMS and operator voice calls). Fig.3 shows how the expansion of Language Translation use cases may happen with the operators in the future.

III. ARCHITECTURE AND EXAMPLES

This section presents the architecture of Nokia speech-to-speech translation solution. After a successful Proof of Concept (PoC) demonstration, the service is already deployed within a European Operator’s network. Fig.4 plots important attributes of the system. Nokia Real-Time Translation server is ready to be connected to Operator’s core NW via a VPN tunnel (both for Circuit-Switched (CS) and for Packet-Switched (PS) core), using SIP and RTP (PCMA) interfaces for call establishment. Nokia Data Center selection depends on Operator’s location, similarly to the selection of Microsoft DC (deployed in US West coast, US East coast, Dublin and Hong-Kong). Secure connectivity (TLS- or IPSec-based) exists between Nokia and Microsoft Azure cloud, the translation service is available via a Websocket connection.

Fig.5 shows more details about the structure of the Operator’s network (including radio access, voice and connection related protocols and network elements), and Nokia Data Center. The connectivity to Operator’s core is the main technical requirement; both MSS (2G and 3G) and IMS (4G and 5G in the future) cores can be connected to Nokia Data Center. Translation can be invoked either from beginning of call or during the call (e.g. based on DTMF). For a detailed description of the interfaces shown in Fig.5, please see ETSI TS 123.002. It is important for the subscribers, that all access types (2G, 3G, fixed, VoLTE, VoWiFi) can be used in the
translation service, i.e. A and B parties can use any device which works in the Operator's network. The two call parties can have even different access technologies (e.g. calling 'A' using 2G, while called 'B' using 3G). Current PoC supports three concurrent calls per operator.

The key idea behind this solution is a beneficial partnering with using Microsoft cloud-based (Azure) machine translation service. Microsoft Translator is built on more than a decade of natural-language research at Microsoft. Microsoft Azure Cognitive API for Translations supports 10 languages for conversational speech translation: Arabic, Chinese (Mandarin), English, French, German, Italian, Japanese, Portuguese, Russian, Spanish, but there are some more in the pipeline. The investment in this direction is continuous, there is a growing feature set with 60 languages for text translation, 20 languages for voice recognition with simple sentences, and Optical Character Recognition (OCR).

Fig. 6 shows how Microsoft Translation works: to translate the "source" speech from one language to a different "target" language. The process goes through a four steps:

1) The first phase involves speech recognition, to convert audio into text. The Automatic Speech Recognition (ASR)-system is optimized for normal conversations.

2) The next phase applies TrueText, which is a Microsoft technology that normalizes the text to make it more appropriate for translation.

3) In the third phase, text is translated by the text translation engine with translation models specially developed for real-life spoken conversations.

4) Finally, text-to-speech conversation happens, when it is necessary to produce the translated audio.

The Speech API towards translation engine of Microsoft is based on two interfaces:

- a RESTful interface, which authenticates and authorizes the usage, the invocation and the control of translation service;
- a WebSocket-based audio interface, which transfers actual speech to and from service.

As Fig. 7 shows, first a Microsoft Azure account registration is needed to obtain a token. The parameters of the connection are the languages, the gender (male/female) and the profanity. It is not possible to detect language or update connection. Two Websockets are needed per conversation, which can be either prefix based or fix. The client application should stream a single channel. The audio stream is encoded with signed 16 bit PCM codec sampled at 16 kHz. The service processes chunk, then provides text and speech. 'A' and 'B' parties hear both the original speech and its translation too. The input is muted when the translation is played. Please note, that the originating MSS and the terminating MSS may be the same network element (MSS orig. and MSS term. in the figure) depending on the call and UE location scenario. This is also true for the MGWs.

Fig. 8 shows that a VoLTE-based solution is also applicable including the Language Translation service. Nokia TAS is responsible for the management of SIP session control and invocation of translation service based on subscription data (HSS). The MRF has interfaces to control translation service (REST) and provide WebSocket interface to transfer audio for translation. Authorization for use of the service is provided by enhancing Advanced Media Services.xml document that is stored in HSS IMS subscription of served user and provided from Nokia TAS to MRF via SIP/Mr' interface. Other parameters may also be defined such as the default language (of the subscriber), the allowed languages or the maximum allowed length of translation, etc.
IV. CONCLUSIONS, VISIONS FOR THE FUTURE

An important question to be asked: how will Operator monetize Language Translation service? Fortunately, many alternatives can be mentioned for collecting the value. First of all, there are some charging related opportunities, like usage-based or time-based charging attached to the call. One price per translated call regardless of call length can be billed or just a fixed price per day / month, or Operator can plan a 'freemium' business model. Furthermore, some ways are opening to get more in-bound roamers — included to SIM-card sales at airports/harbours, and selling more expensive service packages (Language Translation included only to high-end packages). This service can be also included in some physical device; phone or SIM-card or car rental (where B-party pays for translation). A tool can be developed to lock-in enterprise/business customers. In the sales of IMS core, enterprise services can be licensed both from Nokia and service provider point of view, e.g. the translation service for operators that serve enterprise customers such as call center operators.

One option to develop this solution further, is to integrate more translation engines into the solution. This way Nokia Data Center would interface with many selected translation engines, other than Microsoft, so this way providing always the best fit for the Operators’ use cases. Other than voice services, video call subtitling, or SMS translations would also be possible.

ACKNOWLEDGMENT

The authors thank Kalle Luukkanen, Jarkko Pennanen and Ismo Matilainen, all with Nokia Mobile Networks, Finland, for the PoC development and their help in the Budapest demo session. The continuous support of Dr. Gábor Jaró and Toncsi Koczka, both from Nokia Mobile Networks, Hungary is greatly acknowledged.

REFERENCES

[7] ETSI TS 123 002 V14.1.0 (2017-05), Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Network architecture (3GPP TS 23.002, Release 14).


TERMS AND ABBREVIATIONS

1-9 2G 2nd generation mobile network, GSM
2G 3G 3rd generation mobile network, UMTS
3G 4G 4th generation mobile network, LTE
4G A AMR Adaptive Multi-Rate
APR Application Programming Interface
AS Application Server
ASR Automatic Speech Recognition
B B2C Business-to-Consumer
C CJH Consumer-to-Business
CP Control Plane
CS Circuit Switched
CSF Circuit Switched Control Function
D DC Data Center
DTMF Dual-tone multi-frequency signaling
EDGE EDGE Enhanced Datarates for GSM Evolution
EURTAN Evolved-UTRAN
F FW FireWall
GERAN GMS/EDGE Radio Access Network
G GSM Global System for Mobile Communications
GSM (Groupe Spécial Mobile)
GW GateWay
H HSS Home Subscriber Server
I IMS IP Multimedia Subsystem
IN Intelligent Network
IoT Internet of Things
IP Internet Protocol
IPSec IP Security
I LTE Long Term Evolution
N NW Network
M MGW Media GateWay
MRF Media Resource Function
MSC Mobile Switching Center
MSS MSC Server
O OCR Optical Character Recognition
OTT Over-the-Top
P PCM Pulse Code Modulation
PCMA A-law coding, according to ITU-T G.711
PoC Proof of Concept
PS Packet Switched
Q QoS Quality of Service
R RAN Radio Access Network
REST Representational State Transfer
RTP Real-time Transmission Protocol
S SIP Session Initiation Protocol
SMS Short Message Service
SRVCC Single Radio Voice Call Continuity
T TAS Telecommunication Application Server
TC TransCoding
TLS Transport Layer Security
TTS Text-to-Speech
U UE User Equipment
UMTS Universal Mobile Telecommunications System
UP User Plane
UTRAN Universal Terrestrial Radio Access Network
V VoLTE Voice over LTE
VoWiFi Voice over WiFi
VPN Virtual Private Network
W WAV Waveform Audio File Format
WB WideBand
WCDMA Wideband Code Division Multiple Access

Maté Ákos Tündik graduated from the Budapest University of Technology and Economics (BUTE) as a software engineer in 2013. Currently he is a PhD Student of the Laboratory of Speech Acoustics at BUTE Department of Telecommunications and Media Informatics. His main research topics are in connection with speech and language technology: semantic and syntactic analysis of speech recognition output, automatic stress detection and phonological phrasing, punctuation restoration, applying deep learning techniques. He joined Nokia in 2013, as a software engineer. Recently he has been working as a software specialist, and responsible for Telephony Application Server low-level specifications.

Loránd Nagy graduated from the Budapest University of Technology and Economics as a physicist in 2014. Currently he is studying theoretical mathematics also at BUTE. He joined Nokia in 2014, and worked in various R&D positions since. He co-authored technical notes and publications concerning service availability and load distribution in Cloud environments. Currently, he is working as a telecommunication end-to-end network integration professional, specializing in VoLTE network low level design and integration.

Attila Hilt PhD, MSc EE, graduated in electrical engineering from BME, the Technical University of Budapest in 1990. In 1989 he joined TKI, the Research Institute for Telecommunications, Budapest, Hungary. Until 1999, he headed the Telecommunication Test Laboratory of TKI. He worked in the research and development of microwave, millimeter-wave and optical circuits and systems. In May 1999, he received the ‘Doctor of INPG’ degree from the Institute National Polytechnique de Grenoble, France. He received the Ph.D. degree in 2000 from BUTE, the Budapest University of Technology and Economics. He joined Nokia Networks in 2000 where he is working now as Sr. System Specialist. He has been involved in the dimensioning and design of several mobile networks. He participated in the design, testing, piloting, deployment and optimization of several mobile networks, including nationwide modernization projects in Europe. His research interests include measurement techniques, fixed, wireless and mobile communication networks, microwave, millimeter-wave and photonic systems. He authored/co-authored more than 80 papers and more than 120 research, project and test reports, guidelines and specifications. A. Hilt is a member of HTE, the Scientific Association of Infocommunications and MMK, the Hungarian Chamber of Engineers. In 2016 he became a member of the Telecommunications Public Body of the Hungarian Academy of Sciences.

Gergő Bóta graduated from Budapest University of Technology and Economics in 2007 as Electrical Engineer. Right after graduation he joined Nokia as SW engineer and since then has been working in different positions and on different products, starting with Circuit Switched core network elements to WCDMA Radio technologies. He started working in 2016 as an incubation project on the presented Language Translation solution. Recently he has joined Nokia Innovation Steering organization where with his team they are working on many new ideas in the areas of Industry 4.0, Smart Cities, Drone technologies etc.