

**Department of Homeland Security  
(DHS) Science and Technology (S&T)**

**Real-Time Translation**

**TECHNOLOGY SCOUTING  
RESEARCH SUMMARY**

**Date:** August 2019



**Homeland  
Security**

**Science and Technology**

THIS REPORT IS PROVIDED "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. As a condition of the use of this report, the recipient agrees that in no event shall the United States Government, its contractors, or subcontractors be liable for any damages, including, but not limited to, direct, indirect, special, or consequential damages, arising out of, resulting from, or in any way connected to this report or the use of information from this report for any purpose.

# Real-Time Translation



**Overview:** Commercial datasets and open-source research were utilized to compile a list of solution options. A summary of the request is outlined below, and the top identified solutions thus far are displayed on the following pages.

## Problem Description:

When responding to incidents, emergency responders often interact with individuals who speak little or no English, making it very challenging to quickly assess and respond to the problem. In addition, 9-1-1 dispatchers sometimes struggle with understanding non-English speakers when answering emergency calls. Further, there are times when dispatchers or responders cannot determine which language is being spoken or understand what is being said. The inability of on-scene responders and 9-1-1 dispatchers to effectively communicate in real time with non-English speakers or the speech impaired delays response time, which impacts operations when responding to incidents.

## Desired Use Case:

In the desired use case, potential solutions will provide the ability for on-scene emergency responders and 9-1-1 dispatchers to rapidly and effectively communicate with non-English speakers and those unable to communicate verbally. It is anticipated that the solution would be a separate device that will automatically identify, translate and display the language being spoken in real-time during emergency response operations. Responders would carry the device with them on-scene and activate it as soon as communications begin with a non-English speaker or those unable to communicate verbally. The device may be spoken into by all parties involved (e.g., responders, patients, by-standers). Once spoken into, the device would work to automatically identify the language, translate and display the conversation in real-time. It would work at the Public Safety Answering Point (PSAP) for 9-1-1 callers and dispatchers as well as for responders in the field. The translation results should be displayed on a screen for responders and other parties to see so they can understand each other more quickly and accurately.

## Technology Requirements:

The list of the requirements assessed for this report are listed below:

- Detects, identifies and translates various languages in real-time
- Integrates into existing PSAP infrastructure including computer-aided dispatch (CAD), automatic call distributor (ACD) and customer premise equipment (CPE)
- Connects and integrates with legacy architecture
- Is inclusive of text-to-911
- Is cost-effective




Additional desirable capabilities are listed below:

- Connects and integrates with Next-Generation 911 (NG911) architecture
- Displays name of language spoken in real-time

- Translates languages at a 95% accuracy rate
- Includes camera for image capture and translation
- Performs two-way sign language translation for the hearing impaired
- Records audio, video and provides transcription
- Has a pre-record buffer and programmable durations
- Has a graphical user interface (GUI) screen
- Is manually activated
- Be of a size that should not exceed that of a cellular phone
- Includes a keyboard for people who have speech impediments or other disabilities that effect their ability to effectively communicate orally
- Filters out background noise
- Converts voice-to-text during language conversion
- Integrates with other equipment (body-worn cameras, flashlight)
- Processes speech of understand multiple or simultaneous speakers

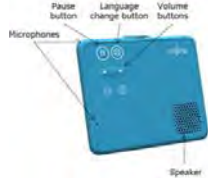

# Real-Time Translation




Solution Options – Physical Solutions				
#	Solution	Description	Requirements	
1	 <p><a href="#">ELSA – Hand-Held Translation Device [Enabling Language Service Everywhere]</a> by RTT Mobile (USA)</p>	<p>The product is a small handheld device that, when activated, connects the user with a network of interpreters who can translate 180 different languages. This allows immediate communication between parties, and the conversations are recorded for future playback. The device contains a high-quality microphone and speaker, enabling clear communication. Many of the network of interpreters have special training in medical, business and government terminology so they can correctly translate even the more complex concepts.</p>	Real-time language detection	Yes
			Real-time language identification	Yes
			Real-time language translation	Yes
			PSAS infrastructure integration	*
			Inclusive of text-to-911	*
			Cost-effective	*
			<i>*More Information Required for Detailed Product Specifications</i>	
2	 <p><a href="#">Foreign Language Translation System [MFLTS]</a> by US Army [Raytheon] (USA)</p>	<p>Consists of two apps, one for real-time, two-way, speech-to-speech translation and one for text-to-text translation of electronic documents, webpages and social media. Initial capability supports translation of two spoken languages (Pashto, Iraqi Arabic), one written language (Modern Standard Arabic) and two language domains (Checkpoint Operations &amp; Base Security). Has been in service since December 2016.</p>	Real-time language detection	Yes
			Real-time language identification	Yes
			Real-time language translation	Yes
			PSAS infrastructure integration	*
			Inclusive of text-to-911	*
			Cost-effective	*
3	 <p><a href="#">SQU.ID SQ.410</a> by Voxtec (USA)</p>	<p>The device is a hands-free translator (originally developed for the US Army) that resembles a rugged walkie-talkie, has a library of over 70 languages to help soldiers and medical personnel communicate with the locals. It is screenless, does not need to connect to a network (which is handy for disaster zones), and does not have to be trained to understand the speaker.</p>	Real-time language detection	Yes
			Real-time language identification	Yes
			Real-time language translation	Yes
			PSAS infrastructure integration	*
			Inclusive of text-to-911	*
			Cost-effective	*

# Real-Time Translation







4	 <p><a href="#">Fujitsu Speech Translation Device</a> by Fujitsu Laboratories (Japan)</p>	<p>An ID card type speech translation device that leaves users' hands completely free. The device, which is worn on the chest, automatically recognizes pauses in conversations and translates conversation content. This device employs various techniques, including omnidirectional microphones, to make it compact and lightweight. Focuses on the "sound channel," which is how the speaker's voice is transmitted to the microphone, using an L-shaped sound channel, to distinguish the target speaker's voice to be translated from sounds from other directions.</p>	Real-time language detection	No
			Real-time language identification	No
			Real-time language translation	Yes
			PSAS infrastructure integration	*
			Inclusive of text-to-911	*
			Cost-effective	Yes
5	 <p><a href="#">Emergency Medical Translator</a> by British Columbia Paramedic (Canada)</p>	<p>A language and symbol-based reference guide designed for medical first responders. The booklet includes: 9 Languages - Questions and statements in English with nine additional translations: Punjabi, Chinese, Korean, Persian (Farsi), Spanish, Arabic, French, Hindi, and Vietnamese; 11 pages of symbols; Language identification guide; Conversions and reference pages.</p> <p><i>Note: This solution is similar to other familiar flip-cards, such as FEMA's 'I SPEAK' cards. Though not new in concept, included for awareness.</i></p>	Real-time language detection	No
			Real-time language identification	No
			Real-time language translation	No
			PSAS infrastructure integration	No
			Inclusive of text-to-911	No
			Cost-effective	Yes

Solution Options – Application Programming Interfaces (API)				
#	Solution	Description	Requirements	
1	 <p><a href="#">Amazon Translate</a> by Amazon Web Services (USA)</p>	<p>Supports the translation of 25 different languages. Its application allows for the performance of both batch translation when there are large quantities of pre-existing text to translate, and real-time translation to deliver on-demand translations of content as a feature of applications. Can instantly translate chat conversations. At time of report, it does not appear that this application allows for speech translation, but future developments of Alexa, the personal assistant technology, may yield improvements to speech recognition/translation capabilities.</p>	Real-time language detection	Yes
			Real-time language identification	Yes
			Real-time language translation	Yes
			PSAS infrastructure integration	*
			Inclusive of text-to-911	*
			Cost-effective	*


# Real-Time Translation



2	 <p><b><u>Dynamic Simultaneous Translation and Anticipation and Controllable Latency (STACL)</u></b> by Baidu (China)</p>	<p>Can begin a translation just a few seconds into a speaker’s speech and finishing seconds after the end of a sentence. It is the opposite of consecutive interpretation, where a translator waits until the speaker pauses to start translating.</p>	<table border="1"> <tbody> <tr><td>Real-time language detection</td><td>Yes</td></tr> <tr><td>Real-time language identification</td><td>Yes</td></tr> <tr><td>Real-time language translation</td><td>Yes</td></tr> <tr><td>PSAS infrastructure integration</td><td>*</td></tr> <tr><td>Inclusive of text-to-911</td><td>*</td></tr> <tr><td>Cost-effective</td><td>*</td></tr> </tbody> </table>	Real-time language detection	Yes	Real-time language identification	Yes	Real-time language translation	Yes	PSAS infrastructure integration	*	Inclusive of text-to-911	*	Cost-effective	*
Real-time language detection	Yes														
Real-time language identification	Yes														
Real-time language translation	Yes														
PSAS infrastructure integration	*														
Inclusive of text-to-911	*														
Cost-effective	*														
3	 <p><b><u>Google Cloud Speech API</u></b> by Google (USA)</p>	<p>API has speech-to- text translation capabilities, has tested well in medical scenarios. Can automatically identify spoken language and return text transcriptions in real time for short-form and long-form audio from multiple audio sources. The API recognizes 120 languages and dialects, including 14 different English dialects, 20 Spanish dialects, and 15 Arabic dialects. Generally available to all third-party developers, Google offers multiple pre-built speech recognition models that are customizable to the use case.</p>	<table border="1"> <tbody> <tr><td>Real-time language detection</td><td>Yes</td></tr> <tr><td>Real-time language identification</td><td>Yes</td></tr> <tr><td>Real-time language translation</td><td>Yes</td></tr> <tr><td>PSAS infrastructure integration</td><td>*</td></tr> <tr><td>Inclusive of text-to-911</td><td>*</td></tr> <tr><td>Cost-effective</td><td>*</td></tr> </tbody> </table>	Real-time language detection	Yes	Real-time language identification	Yes	Real-time language translation	Yes	PSAS infrastructure integration	*	Inclusive of text-to-911	*	Cost-effective	*
Real-time language detection	Yes														
Real-time language identification	Yes														
Real-time language translation	Yes														
PSAS infrastructure integration	*														
Inclusive of text-to-911	*														
Cost-effective	*														
4	 <p><b><u>Babel Program (IARPA)</u></b> by IARPA (USA)</p>	<p>Developing speech recognition technology that can be rapidly applied to any human language. Will be able to generate a speech transcription system for any new language within one week to support keyword search performance for effective triage of massive amounts of speech recorded in real-world situations.</p>	<table border="1"> <tbody> <tr><td>Real-time language detection</td><td>Yes</td></tr> <tr><td>Real-time language identification</td><td>Yes</td></tr> <tr><td>Real-time language translation</td><td>Yes</td></tr> <tr><td>PSAS infrastructure integration</td><td>*</td></tr> <tr><td>Inclusive of text-to-911</td><td>*</td></tr> <tr><td>Cost-effective</td><td>Yes</td></tr> </tbody> </table>	Real-time language detection	Yes	Real-time language identification	Yes	Real-time language translation	Yes	PSAS infrastructure integration	*	Inclusive of text-to-911	*	Cost-effective	Yes
Real-time language detection	Yes														
Real-time language identification	Yes														
Real-time language translation	Yes														
PSAS infrastructure integration	*														
Inclusive of text-to-911	*														
Cost-effective	Yes														
5	 <p><b><u>IBM Watson Speech-to-Text API</u></b> by IBM (USA)</p>	<p>Can automatically transcribe audio from 7 languages, dialects excluded, in real-time; it can do so even with lower quality recordings, conversations with multiple speakers, and more. The model is customizable to recognize certain names, sensitive subjects, names of individuals, and other content that is relevant to the use case. Can identify and determine repeated words and tones in conversations, recognizing patterns and offering analytic insights on the audio content.</p>	<table border="1"> <tbody> <tr><td>Real-time language detection</td><td>Yes</td></tr> <tr><td>Real-time language identification</td><td>Yes</td></tr> <tr><td>Real-time language translation</td><td>Yes</td></tr> <tr><td>PSAS infrastructure integration</td><td>*</td></tr> <tr><td>Inclusive of text-to-911</td><td>*</td></tr> <tr><td>Cost-effective</td><td>*</td></tr> </tbody> </table>	Real-time language detection	Yes	Real-time language identification	Yes	Real-time language translation	Yes	PSAS infrastructure integration	*	Inclusive of text-to-911	*	Cost-effective	*
Real-time language detection	Yes														
Real-time language identification	Yes														
Real-time language translation	Yes														
PSAS infrastructure integration	*														
Inclusive of text-to-911	*														
Cost-effective	*														



# Real-Time Translation



6	 <p><a href="#">Speech Translator API</a> by Microsoft Azure (USA)</p>	<p>A cloud-based automatic translation service which enables developers to add end-to-end, real-time, speech translations to their applications or services. The Translator supports more than 60 languages for text translation through the Microsoft Translator API.</p>	Real-time language detection	Yes
			Real-time language identification	Yes
			Real-time language translation	Yes
			PSAS infrastructure integration	*
			Inclusive of text-to-911	*
			Cost-effective	*


## Additional Consumer-focused Solutions

*There are several consumer-grade, handheld translation devices designed and marketed primarily toward travelers. While these devices do not meet many of the requirements listed in the SOO, some solutions are included below for awareness of devices that can address a simpler use case.*

#	Solution	Description
1	 <p><a href="#">ili</a> by Logbar (USA, Japan)</p>	<p>A wearable translator. A device developed for travelers and works without being connected to the Internet. It currently translates between English and Japanese, Mandarin, Spanish, and Korean.</p>
2	 <p><a href="#">Ambassador</a> by Waverly Labs (USA)</p>	<p>Over-the-ear design. Seamless pairing, simple sharing and adaptable settings for live translations across the globe.</p> <p>Features:</p> <ul style="list-style-type: none"> <li>• Smooth sharing with guests by pairing multiple Ambassadors to one smartphone</li> <li>• Over-the-ear design for sharing, longer battery life, and signal quality</li> <li>• Advanced microphone array</li> <li>• Translate and understand dozens of languages / dialects with guests and groups</li> </ul>

# Real-Time Translation



3	 <p><a href="#">Enence</a> by Muama (USA)</p>	<p>A smart voice translator, a real time speech interactive translator that can translate 40 languages such as English, Chinese, Japanese, Spanish and more. Using intelligent speech recognition technology, it can translate recorded sound into text, and simultaneously transmit into voice output.</p>
---	--	---