Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language 27 (2013) 528–553
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Evaluation methodology and metrics employed to assess the TRANSTAC two-way, speech-to-speech translation systems☆

Gregory A. Sanders[a,*], Brian A. Weiss[a], Craig Schlenoff[a], Michelle P. Steves[a], Sherri Condon[b]

[a] National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[b] The MITRE Corporation, 7515 Colshire Drive, Mailstop H305, McLean, VA 22102, USA

Received 24 February 2010; received in revised form 28 April 2011; accepted 3 May 2011
Available online 7 May 2011

Abstract

One of the most difficult challenges that military personnel face when operating in foreign countries is clear and successful communication with the local population. To address this issue, the Defense Advanced Research Projects Agency (DARPA) is funding academic institutions and industrial organizations through the Spoken Language Communication and Translation System for Tactical Use (TRANSTAC) program to develop practical machine translation systems. The goal of the TRANSTAC program is to demonstrate capabilities to rapidly develop and field free-form, two-way, speech-to-speech translation systems that enable speakers of different languages to communicate with one another in real-world tactical situations without an interpreter. Evaluations of these technologies are a significant part of the program and DARPA has asked the National Institute of Standards and Technology (NIST) to lead this effort. This article presents the experimental design of the TRANSTAC evaluations and the metrics, both quantitative and qualitative, that were used to comprehensively assess the systems' performance.

Keywords: Evaluation; Machine translation; NIST; Performance metrics; Bidirectional speech-to-speech translation; TRANSTAC

1. Overview

[1]The Spoken Language Communication and Translation System for Tactical Use (TRANSTAC) program is a Defense Advanced Research Projects Agency[2] (DARPA) advanced technology research and development program. Typically, U.S. Soldiers and Marines speak only English, yet are deployed to countries where the local civilian populations typically speak only other languages. The goal of the TRANSTAC program is to demonstrate capabilities to rapidly develop and field free-form, two-way (both from English and to English) speech-to-speech machine translation

☆ This paper has been recommended for acceptance by 'Shrikanth Narayanan'.

* Corresponding author. Tel.: +1 301 975 4451; fax: +1 301 670 0939.
E-mail addresses: gregory.sanders@nist.gov (G.A. Sanders), brian.weiss@nist.gov (B.A. Weiss), craig.schlenoff@nist.gov (C. Schlenoff), msteves@nist.gov (M.P. Steves), scondon@mitre.org (S. Condon).

[1] DARPA Clearance: Approved for public release, distribution unlimited.

[2] The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

0885-2308/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.csl.2011.05.001

Fig. 1. TRANSTAC technology form factors.

systems that enable speakers of different languages, particularly local civilians and U.S. military personnel, to communicate with one another in real-world tactical situations without an interpreter. Under the TRANSTAC program, several prototype systems have been developed for domains including force protection, civil affairs, and medical screening, in languages including Iraqi Arabic (IA), Farsi, Pashto, and Dari Systems have been demonstrated on a wide range of platforms with varying performance.

The primary use cases involve spoken communication between US military personnel and local foreign language speakers. Though never expected to replace human translators, the TRANSTAC systems will help when there is a shortage of translators in key foreign languages and will aid in quickly providing translation capabilities when languages unexpectedly become important, as may happen for a sudden, unexpected military deployment or after a natural disaster.

The term *speech-to-speech machine translation system* means that both the input and the output are spoken language – the users speak into the system's microphone and the system speaks out the translation. When speaking of translation, we call the input language the *source language* and we call the output language the *target language*. For example, when translating from Arabic to English, Arabic is the source language and English is the target language. When translating in the opposite direction, English would be the source and Arabic the target language.

All the TRANSTAC spoken language translation systems consist of the following three components. First, when the user speaks into the system's microphone, the Automatic Speech Recognition (ASR) component of the TRANSTAC system analyzes the speech and generates a source-language textual transcription of what was said. Next, the Machine Translation (MT) component of the TRANSTAC system translates that source-language transcription to text in the target language. Finally, the Text-To-Speech (TTS) component of the TRANSTAC system converts the target language text into an audio file, which is then output to the target language speaker.

This same process happens in both directions: translating what the English speaker says (typically a question) into the foreign language and translating what the foreign language speaker says (typically an answer) into English.

Multiple research teams have participated in the TRANSTAC program. The form factor of the systems varied among the teams. Some chose to focus on platforms with more memory and processing power rather than ergonomics, and used laptops. Others opted to trade-off processing power in favor of employing more wearable systems such as the Lynx[3] platform or the Panasonic U1. The interface to the system also varied by team. Some used hand-held microphones, some used head-mounted microphones and some used a platform-embedded microphone (e.g., the microphone built into the Lynx). Some examples of these hardware platforms are shown in Fig. 1.

The National Institute of Standards and Technology (NIST) led the Independent Evaluation Team (IET) for the TRANSTAC effort, with the goal of systematically analyzing the performance of the TRANSTAC systems periodically throughout their continuing development (Weiss et al., 2008; Schlenoff et al., 2009). NIST implemented a multi-faceted

[3] Certain commercial equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedures adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the equipment, instruments, software, or materials are necessarily the best available for the purpose.