AL-FARABI KAZAKH NATIONAL UNIVERSITY
International Relations Department
Chair of Diplomatic Translation
Translation business in the field of international and legal relations
"Simultaneous Translation of Monologue and Dialogical Speech"
2022-2023 academic year fall semester

Lecture 4

Module 1: Tenets of Translation Studies

Lecture 4: Simultaneous interpretation of pre-translated discourse

Plan of the lecture

1. Introduction
2. Simultaneous translation of lectures and speeches
3. Conclusion
4. References

Aspects of the lecture

1. Spoken language translation
2. Vocabulary, dictionary, and language model
3. Machine translation

Goals of the lecture

1. Explain the acoustic model of SI
2. Introduce speech recognition techniques
3. Familiarize with the system of spoken language translation
4. Highlight the importance of coordination and presentation skills

Basic concepts

Speech-to-speech translation, spoken language translation, machine translation, speech recognition, lecture recognition, lectures, speeches, machine translation, and etc.

With advancing globalization, effective cross-cultural communication is rapidly becoming an essential part of modern life. Yet, an estimated 4,000 to 6,000 languages still separate the people of the world, and only a small intellectual elite gather by agreeing to communicate in a common language, usually English. This situation creates a dilemma between a need to integrate and interact and the need to maintain cultural diversity and integrity. It also limits international

discourse to an adopted communication medium that does not always reflect the culturally nuanced, individual subtleties of the global human experience. To provide access to other languages unimpeded, however, requires translation, but due to the enormous cost of human translation, only a small fraction of text documents are presently translated and only a handful of human spoken encounters are actually interpreted, if they even take place at all as a result of the separation.

All this could change if affordable, real-time, simultaneous speech-to-speech translation (SST) was possible by machine. Machines never suffer fatigue, nor are they limited in principle by memory size or processing speed. They could potentially also open communication channels between uncommon spoken languages, so as to broaden the exchange and interaction between the people of the world.

As a result, automatic SST as a research endeavor has been enjoying increasing attention. After humble first beginnings in the 1990s, when limited phrase-based and first domain-dependent speech translation systems were proposed (Waibel et al. 1991; Morimoto et al. 1993) large-scale domain-unlimited translation systems are now beginning to emerge (Waibel and Fügen 2008). Three recent projects in particular have begun to advance the state of the art from domain-dependent to domain-independent capabilities: the NSF-ITR project STR-DUST, the EC-IP project TC-STAR, and the DARPA project GALE (see Sect. 3 for more details about these projects). They bring within reach the translation of broadcast news and political speeches between several major languages.

As systems developed within these three projects focus on the offline processing of huge data material, translation quality instead of speed was of primary importance. As we will show in this paper, however, even simultaneous domain-unlimited translation is possible thereby opening up the range of applications to interactive lectures, seminars, speeches, and meetings. To do so, we have combined advances in speech translation technology from these projects with advances in human interface technology developed under the European Commission (EC) integrated project CHIL (Computers in the Human Interaction Loop) (Waibel et al. 2004; Waibel and Stiefelhagen 2009). While the work within CHIL focused on the delivery and interface aspect of a simultaneous translation service and the integration into the Smart Room, core recognition and translation components were advanced under TC-STAR. In the present paper, we describe how these technologies were extended to deliver real-time simultaneous speech translation of spontaneous lectures. We will discuss how the added requirements found in technical university lectures and seminars, such as real-time performance, online sentence-level segmentation, special terms for lecture vocabularies, and disfluencies, can be addressed. In human user experimentation, finally, we will attempt to establish an assessment of whether simultaneous lecture translation Although

the terms "translation" and "interpreting" are used interchangeably in everyday speech, they vary greatly in meaning. Both refer to the transfer of meaning between two languages. However, translation refers to the transfer of meaning from text to text, often with time and access to resources such as dictionaries, glossaries, etc. On the other hand, interpreting consists of facilitating oral or sign language communication, either simultaneously or consecutively, between two or more speakers who are not speaking the same language.1 The profession expects interpreters to be more than 80% accurate and translations, by contrast, over 99% accurate.2 Simultaneous interpreting is sometimes incorrectly referred to as "simultaneous translation" and the interpreter as the "translator". However, in computer science the term "machine translation" (MT) is commonly used for systems translating text or speech from one language to another. The reason for this is that in the past the main focus of MT was the translation of text and only recently is SST attracting a wider interest. Therefore, throughout this paper we use the term translation as in "SST", "simultaneous translation" or "simultaneous speech translation" to mean the automatic interpretation of spoken language.

Follow-up questions

1. Name main characteristics of acoustic model of SI
2. Speak about speech recognition techniques
3. Describe In-conference knowledge acquisition in the context of Communicative Situation
4. Speak about meeting formats

References

1. Doddington G (2002) Automatic evaluation of MT quality using n-gram co-occurrence statistics. In: Proceedings of human language technology conference 2002, San Diego, CA, 138-145

2. Eide E, Gish H (1996) A paramteric approach to vocal tract length normalization. In: 1996 IEEE international conference on acoustics, speech, and signal processing, Atlanta, Georgia, pp 346–348

3. Finke M, Geutner P, Hild H, Kemp T, Ries K, Westphal M (1997) The Karlsruhe-verbmobil speech recognition engine. In: 1997 IEEE international conference on acoustics, speech, and signal processing (ICASSP'97), Munich, Germany, pp 83–86