Effect of Machine Translation in Interlingual Conversation
Kotaro Hara and Shamsi Iqbal
こんにちは！
こんにちは！
Hello!
Successful intercultural communication is important in:

- Education
- Business
- Health-care

Yet, a language barrier can be a challenge.
Simultaneous interpretation is **only for privileged few**
"The infrequent use of interpreters* in the delivery ward was among the most important reasons for the reduced quality of [health] care."

Kale and Syed (2010)

* Interpreter: a person who translates the words that someone is speaking into a different language (Merriam-Webster)
Automatic Spoken Language Translation
OPTIMIZING SENTENCE SEGMENTATION FOR SPEECH TRANSLATION

Shenhar Eloe, Ben Lin, Tom Schatz


tact\tLanguage Technology Institute, University of Pennsylvania, Philadelphia, PA, 19174

Abstract

This paper proposes a novel approach to automatic speech translation that is based on a statistical language model. The approach is based on the idea that, for each sentence in the source language, a probability distribution over the possible translations is estimated. This distribution is then used to select the most likely translation for the sentence.

1. Introduction

There is a growing interest in automatic speech translation systems for a variety of applications, including machine translation, document translation, and speech-to-speech translation. In this paper, we focus on the development of a system that can automatically translate speech from one language to another.

2. Related Work

Many previous approaches to automatic speech translation have been proposed, including rule-based systems, statistical models, and neural networks. In this paper, we evaluate a system that is based on a statistical language model.

3. Methodology

The system is based on a trigram language model, which is trained on a large corpus of speech data.

4. Results

The system achieves a translation accuracy of 85.5%.

5. Conclusion

The results of this study show that automatic speech translation is a feasible technology that can be used in a variety of applications.

Development of a Simultaneous Interpretation System for Face-to-Face Services and Its Evaluation Experiment in Real Situation

Akiho Sakamoto, Ken'ichi Watanabe, Satoshi Kamataki and Kazuo Sumita

Knowledge Media Laboratory, Corporate Research & Development Center, Toshiba Corporation
akikii@kanata.com.jp

Abstract

We developed a simultaneous interpretation system for face-to-face services at shops, farms, and so forth. The system supports interpretation between Japanese, English, or Japanese and Chinese. In practice, it proves users' continuous and spontaneous speech and automatically produces interpretation results. We conducted a field test of the system with an actual user to evaluate the system's effectiveness.

Mobile Speech-to-Speech Translation of Spontaneous Dialog: An Overview of the Final Verbmobil System

Wolfgang Wahlster

Deutsche Telekom AG, Heidelberg, Germany

Abstract

Verbmobil is a product that provides mobile phone users with simultaneous speech-to-speech translation services. In this paper, we present an overview of the technology and the system architecture of Verbmobil, focusing on the key components and their interrelations. We also discuss the challenges and future directions for this technology.

1. Introduction

Verbmobil is a mobile phone application that provides instant speech-to-speech translation services. The aim of the system is to enable cross-linguistic communication by providing real-time translation of spoken speech.

2. Related Work

There is a growing interest in research on speech-to-speech translation, and various approaches have been proposed, including rule-based systems, statistical models, and neural networks.

3. Methodology

The system is based on a trigram language model, which is trained on a large corpus of speech data.

4. Results

The system achieves a translation accuracy of 85.5%.

5. Conclusion

The results of this study show that automatic speech translation is a feasible technology that can be used in a variety of applications.

Verbmobil is a product that provides mobile phone users with simultaneous speech-to-speech translation services. In this paper, we present an overview of the technology and the system architecture of Verbmobil, focusing on the key components and their interrelations. We also discuss the challenges and future directions for this technology.

1. Introduction

Verbmobil is a mobile phone application that provides instant speech-to-speech translation services. The aim of the system is to enable cross-linguistic communication by providing real-time translation of spoken speech.

2. Related Work

There is a growing interest in research on speech-to-speech translation, and various approaches have been proposed, including rule-based systems, statistical models, and neural networks.

3. Methodology

The system is based on a trigram language model, which is trained on a large corpus of speech data.

4. Results

The system achieves a translation accuracy of 85.5%.

5. Conclusion

The results of this study show that automatic speech translation is a feasible technology that can be used in a variety of applications.

Verbmobil is a product that provides mobile phone users with simultaneous speech-to-speech translation services. In this paper, we present an overview of the technology and the system architecture of Verbmobil, focusing on the key components and their interrelations. We also discuss the challenges and future directions for this technology.

1. Introduction

Verbmobil is a mobile phone application that provides instant speech-to-speech translation services. The aim of the system is to enable cross-linguistic communication by providing real-time translation of spoken speech.

2. Related Work

There is a growing interest in research on speech-to-speech translation, and various approaches have been proposed, including rule-based systems, statistical models, and neural networks.

3. Methodology

The system is based on a trigram language model, which is trained on a large corpus of speech data.

4. Results

The system achieves a translation accuracy of 85.5%.

5. Conclusion

The results of this study show that automatic speech translation is a feasible technology that can be used in a variety of applications.

Verbmobil is a product that provides mobile phone users with simultaneous speech-to-speech translation services. In this paper, we present an overview of the technology and the system architecture of Verbmobil, focusing on the key components and their interrelations. We also discuss the challenges and future directions for this technology.

1. Introduction

Verbmobil is a mobile phone application that provides instant speech-to-speech translation services. The aim of the system is to enable cross-linguistic communication by providing real-time translation of spoken speech.

2. Related Work

There is a growing interest in research on speech-to-speech translation, and various approaches have been proposed, including rule-based systems, statistical models, and neural networks.

3. Methodology

The system is based on a trigram language model, which is trained on a large corpus of speech data.

4. Results

The system achieves a translation accuracy of 85.5%.

5. Conclusion

The results of this study show that automatic speech translation is a feasible technology that can be used in a variety of applications.
Fügen et al. (2007) noted that “automatic systems can already provide usable information for people.”
Little research has paid attention to how people interact with spoken language translation technology.
Very little knowledge about Spoken Language Translation in HCI
The void that we are trying to fill

Very little knowledge about Spoken Language Translation in HCI
Background
背景
Hamon O., et al., End-to-End Evaluation in Simultaneous Translation, *ECACL2009*
Hello! こんにちは

Spoken Language Translation System

Automatic speech recognition

Machine translation

Speech synthesis

Speech

Text

Hamon O., et al., End-to-End Evaluation in Simultaneous Translation, ECACL2009
Danger.
Do not enter

абуないから、はいってはいけません
Because you are dangerous,
you must not enter.

Effect of machine translation on text-based communication
Effect of speech recognition on interlingual conversation

Gao et al. CHI2014
Key Point

Speech Recognition

Machine Translation

Text Output
Key Point

Effect of spoken language translation system to interlingual communication is under-explored.
Evaluation of NESPOLE!
Constantini et al. 2002

• Examined usability of spoken language translation system

• No detailed discussion on how people adapted in using the system

• Push-to-talk interface made it hard to analyze turn-taking behavior
Goal

Explore how spoken language translation affects a natural conversation between people speaking different languages.
Video to show how the system works:
One side speaks in English, another side in German.

Skype Translator Demo, English-German Conversation
Video to show how the system works:
One side speaks in English, another side in German.
Video chat interface
Video to show how the system works:
One side speaks in English, another side in German

Closed Caption (CC)
One side speaks in English, another side in German.

Translated Text-to-Speech (TTS) Audio
Video to show how the system works:
One side speaks in English, another side in German.

Translator Tool | Interface Components

Headset
Microphone
Translator Tool

Quantitative Analysis

Study Method

Content Analysis
<table>
<thead>
<tr>
<th>Study Method</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 French-German pairs (N=16 participants)</td>
<td><img src="image" alt="French-German pairs" /></td>
</tr>
<tr>
<td>15 English-German pairs (N=30 participants)</td>
<td><img src="image" alt="English-German pairs" /></td>
</tr>
</tbody>
</table>
Study Method | Participants

<table>
<thead>
<tr>
<th>No common language</th>
<th>With common language (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 French-German pairs (N=16 participants)</td>
<td>15 English-German pairs (N=30 participants)</td>
</tr>
</tbody>
</table>

All participants could speak English reasonably well.
I will mainly talk about this study.

- 8 French-German pairs (N=16 participants)
- 15 English-German pairs (N=30 participants)

With common language (English)

No common language
Olivia trainierte für den Tanzwettbewerb.
(Olivia was practicing for the dance-off)
A starting sentence was provided to the participant.

Olivia trainierte für den Tanzwettbewerb.
(Olivia was practicing for the dance-off)

Translate

A **starting sentence** was provided to the participant.
<table>
<thead>
<tr>
<th>Study Method</th>
<th>Conversation Task</th>
</tr>
</thead>
</table>

**Conversation Task**

<table>
<thead>
<tr>
<th>German Speaker</th>
<th>French Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olivia trainierte für den Tanzwettbewerb. (Olivia was practicing for the dance-off)</td>
<td>Je ne savais pas Olivia dansé. Combien de temps elle s'y adonne ? (I didn’t know Olivia danced. How long has she been practicing?)</td>
</tr>
</tbody>
</table>
Olivia trainierte für den Tanzwettbewerb. (Olivia was practicing for the dance-off)

Seit sechs Jahren (For six years)

Je ne savais pas Olivia dansé. Combien de temps elle s'y adonne ? (I didn’t know Olivia danced. How long has she been practicing?)

C'est une longue période. (That's a long time.)
Conversation Task

The goal of each task was to **collaboratively construct a coherent story**

We asked participants to perform **9 conversation tasks (~3 min ea.) in their respective languages** and an additional conversation task in English.

Tasks were conducted using **three different settings of the translator tool**
Study Method | Interface Settings

Interface Settings

Closed Caption & Text-to-Speech (CC & TTS)  Text-to-Speech only (TTS)

Closed Caption only (CC)
With Closed Caption and Text-to-Speech
yeah. she was a sucking on a lollipops shoes walking along.
the translation was a little strange the he has not enough sleep had had
the translation was a little strange he has not enough sleep had had
die übersetzung war ein bisschen komisch aus ihm hat er nicht genug geschlafen gehabt.
Interface Settings

Round 1

Closed Caption and Text-to-Speech

Closed Caption

Text-to-Speech

Order was permuted for each pair for counter balancing
9 tasks with different interface settings and different starting sentences
Interface Settings

Round 1  |  Round 2  |  Round 3

CC  |  CC  |  CC

Eng  |  CC  |  CC

A baseline English task to compare with/without translation settings
Data

Round 1

Post-task survey (after every task)

E.g., “We had a successful conversation.”
## Data

### Round 1

<table>
<thead>
<tr>
<th>Study Method</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-round survey (after every 3 tasks)</td>
<td>Interface setting preference ranking</td>
</tr>
<tr>
<td>CC &amp; TTS: Best, CC: Neutral, TTS: Worst</td>
<td></td>
</tr>
</tbody>
</table>

### Round 3

Eng
Study Method | Data

Data

Round 1

Round 2

Round 3

Eng  

Post-session survey and Interview
• Overall interface setting preferences

• Whether people became used to using the translator tool
Analysis Method

Three interface settings — With-in subject

Three rounds (nine tasks) — With-in subject

Two languages — Between subject

3 x 3 x 2 mixed design study
Analysis Method

Transformed ordinal data in survey responses with aligned rank transformation

We analyzed the data with restricted maximum likelihood model
Interface Preference Results

Which interface setting did you favor the most and least?

Closed Caption & Text-to-Speech  Closed Caption only  Text-to-Speech only
Quantitative Analysis | Interface Preference Results

Most Favored Interface Setting by French-German Group

Percentage of people who favored the interface setting

- Closed Caption & Text-to-Speech: 59%
- Closed Caption only: 35%
- Text-to-Speech only: 0%
Quantitative Analysis | Interface Preference Results

Most Favored Interface Setting by French-German Group

- Interface settings with closed caption were preferred

- 59% for Closed Caption & Text-to-Speech
- 35% for Closed Caption only
- 6% for Text-to-Speech only
Quantitative Analysis | Interface Preference Results

Most Favored Interface Setting by French-German Group

- **Closed Caption & Text-to-Speech**: 59%
- **Closed Caption only**: 35%
- **Text-to-Speech only**: 6%

24% more people preferred to have text-to-speech.
Quantitative Analysis | Interface Preference Results

Most Favored Interface Setting by French-German Group

Set the Closed Caption & Text-to-Speech to the default, but allow users to turn on/off setting.
Perceived Conversation Quality over Rounds

5-point Likert scale rating (higher is better)

- Round 1: 2.2
- Round 2: 2.2
- Round 3: 2.7

Average 5-point Likert scale rating: 2.4
Perceived Conversation Quality over Rounds

5-point Likert scale score (higher is better)

Overall Conversation Quality

Participants felt that their conversation quality improved

$F_{2,112} = 6.275, \ p < 0.01$; Error bars are standard errors
Perceived Conversation Quality over Rounds

5-point Likert scale score (higher is better)

Perceived conversation quality is not as good as that of English task.
Translator Tool

Quantitative Analysis

Skype translator

Study Method

Content Analysis
• Why people were satisfied or dissatisfied with the experience
• How people adapted to using the translator tool
Method

Post-session interviews were transcribed by authors.

Interview transcripts and survey responses were coded with a content analysis method; recurring themes were extracted and grouped together.
62.5% of the participants noted that proper noun recognition errors hindered the conversation.
'Ava', there was no way it was getting ‘Ava’ no matter how many times I tried. So in German, when you say ‘but,’ it sounds very similar. So that’s what it was picking up.

French-German Pair 4, German speaker
‘Ava’, there was no way it was getting ‘Ava’ no matter how many times I tried. So in German, when you say ‘but,’ it sounds very similar. So that’s what it was picking up.

Lack of an adaptation technique

French-German Pair 4, German speaker
of the participants noted that grammatical errors and misplacement of words were problematic.
Because [...] German sentence structure is so different, [translated] sentence wouldn’t make any sense when [synthesized speech] said it, but when you see it on the screen, you can kind of reorganize the words in the way it supposed to go. And then it makes sense.

French-German Pair 3, German speaker
Because [...] German sentence structure is so different [translated] sentence wouldn’t make any sense when [synthesized speech] said it, but when you see it on the screen, you can kind of reorganize the words in the way it supposed to go. And then it makes sense.
31.3% of the participants noted that original speech was useful.
[Having partner’s original voice] humanizes the relationships as well, because if you only get the robotic voice, I think it just cuts you off from the person you are engaged with. Whereas you talk, then it feels much more human relationship.
31.3% of the participants mentioned that they forgave the errors of the translator tool.
My sentence became shorter, I pronounced more clearly, I was speaking ... I mean, if I were speaking to someone who’s not a native German, and if I didn’t have a translator, then I would also slow down, obviously.

French-German Pair 5, German speaker
Difference in Language Groups

French-German Group

English-German Group
My partner and I would begin to speak then the translation from the previous statement would catch up, translation and speaker would then all be talking at the same time.

English-German Pair 4, English speaker
Speech Overlap

25.0% of the participants mentioned that they *speech overlap* was a problem.

French-German Group
Participants from the English-German group perceived speech overlap as a more severe problem.

This was partly because German speakers could respond without seeing/hearing translation.
Design Guidelines

Section Summary
The analysis suggests that speech recognition errors, translation errors, and difficulties in turn taking decreased the usability of the system. To deal with the imperfect system, participants adjusted the way that they speak, and system designers added adaptations to compensate for these issues. Some problems were more severe because of the lack of adaptation techniques for some translation scenarios. Design guidelines to mitigate these problems are presented.

DESIGN GUIDELINES
We discuss design guidelines that emerged as a result of the analysis. Though they are not meant to be exhaustive, we believe they cover fundamental requirements for designing typical SLT systems.

Support Utensils Adaptation for Speech Production
It is unlikely that MT and ASR will perform perfectly in the near future. However, system designers can help improve user experience by designing their application to support users' adaptation for speech production. There is more adaptation for speech production when the MT application’s errors are more frequent, and participants’ adaptation mechanisms for users know what translation is unsuccessful to allow them to effectively repeat and rephrase their utterances. Participants found closer caption useful for identifying speech recognition errors and subsequently adjusting the way they speak. Nevertheless, because there were no clear indications of why the recognition failed, participants tried various adaptation techniques. Typical speech recognition software produces a confidence score for transcription accuracy and information is presented along with other information. Presenting why it had low confidence to the user allows them to better adjust the way they speak. For example, a user can slow down their speech if the system shows low speech recognition confidence with fast utterance speed.

Support Fallback Strategies with Non-verbal Input
We found that some speech recognition problems were more severe due to the lack of effective correction techniques. System designers should consider offering alternative input methods to provide fallback options for problems such as proper word recognition and statement recognition classification. Allowing editable text-based input on top of speech input would mitigate the problem with ASR’s out-of-vocabulary problem.

Support Comprehension of Messages
Closed captions helped participants comprehend imperfect translation, however, some were frustrated with it not persisting on a screen. Providing a history of translation on a screen could help users to comprehend a long message. Interface designers should consider how to show old transcriptions/translations on limited display of real estate.

Participants mentioned that their partners’ original speech could be used to extract information that during translation. For example, it allows one to pick up a proper noun in a sentence and convey a spoken feeling in it. At the same time, some people may find it redundant. As there would be diverse preferences for hearing/not hearing original speech, we recommend system designers to allow users to easily adjust the volume of original speech.

Support Users’ Turn Taking
Increasing awareness of the system state would benefit turn taking. People can use information about what the system is currently doing to determine when they can safely speak and avoid speech overlap with the system. Showcasing how recognition and translation are occurring is also useful as users can see if the system is pausing pronunciation to the machine translation component or not.

Preference for having or not having TTS was polarized; some preferred having it to gain additional information, and others mentioned that it disrupted turn taking. To accommodate participants with different preferences, we suggest designing to make it easier to turn TTS on and off. Other potential design directions to reduce turn taking costs would be: (i) make TTS smarter to avoid it speaking while a user is speaking; (ii) simplify isolating TTS from the original users’ voices (e.g., ducking the volume of original voices, 3D audio effect to simulate different audio source position); (iii) allow users to abort ASR and MT to avoid immediate translation due to speech collision.

LIMITATIONS AND FUTURE WORK
Our German participants in the English group had reasonable English proficiency. This may have impacted their experience with the tool as they were able to comprehend what their partner said without the translation. However, we feel that analyzing this population is important as some of our potential users could be those who understand the other language, but prefer speaking in their own language.
Design Guidelines

- Support Users’ Adaptation for Speech Production
- Offer Fallback Strategies with Non-verbal Input
- Support Comprehension of Messages
- Support Users’ Turn Taking
Limitation | German Speakers’ English Proficiency
Limitation | Baseline Task

 Interpreter

![skype translator logo](skype_translator.png)
Future Work

Analysis of conversation

Conduct follow up summative studies
Summary

We used Skype Translator to **elicit interaction problems in using spoken language translation** for interlingual conversation.

Our analysis revealed **how people adjusted** (or **could not adjust**) their behaviors to overcome the problems.
Microsoft Skype Translator team
Questions?

@kotarohara_en
Image and Icon Credit

- Friendlies by Mo Riza (https://www.flickr.com/photos/moriza/2565606353)
- 133/2011 euroscola by rosipaw (https://www.flickr.com/photos/rosipaw/5768035647/)
- Japantimes (http://www.japantimes.co.jp/news/2014/05/08/world/social-issues-world/births-fall-economies-may-falter)
- Windows8core.com (http://www.windows8core.com/bing-translator-app-for-windows-8rt-lands-in-store-review/)
- Jigsaw puzzle by James Petts
- Speech bubble from Wikimedia Commons (http://commons.wikimedia.org/wiki/File:Speech_bubble.svg)
- Head silhouette from Channelweb (http://www.channelweb.co.uk/crn-uk/news/2244637/good-week-bad-week)
- Microphone from Uxrepo.com (http://uxrepo.com/icon/mic-by-mfg-labs)
- Funny Signs from Fanpop (http://www.fanpop.com/clubs/picks/images/1771904/title/funny-signs-photo)
- Compass Zoomed by Gwgs (https://www.flickr.com/photos/gwgs/346806882)
- Beaker by Emily van den Heever (https://thenounproject.com/EmvdHeever/)