

# Speech translation in human-to-human interaction: Skype Translator

InterAct25 – Baden-Baden

July 14, 2016

Chris.Wendt@microsoft.com

@Tian500

Group Program Manager – Machine Translation – Microsoft Research

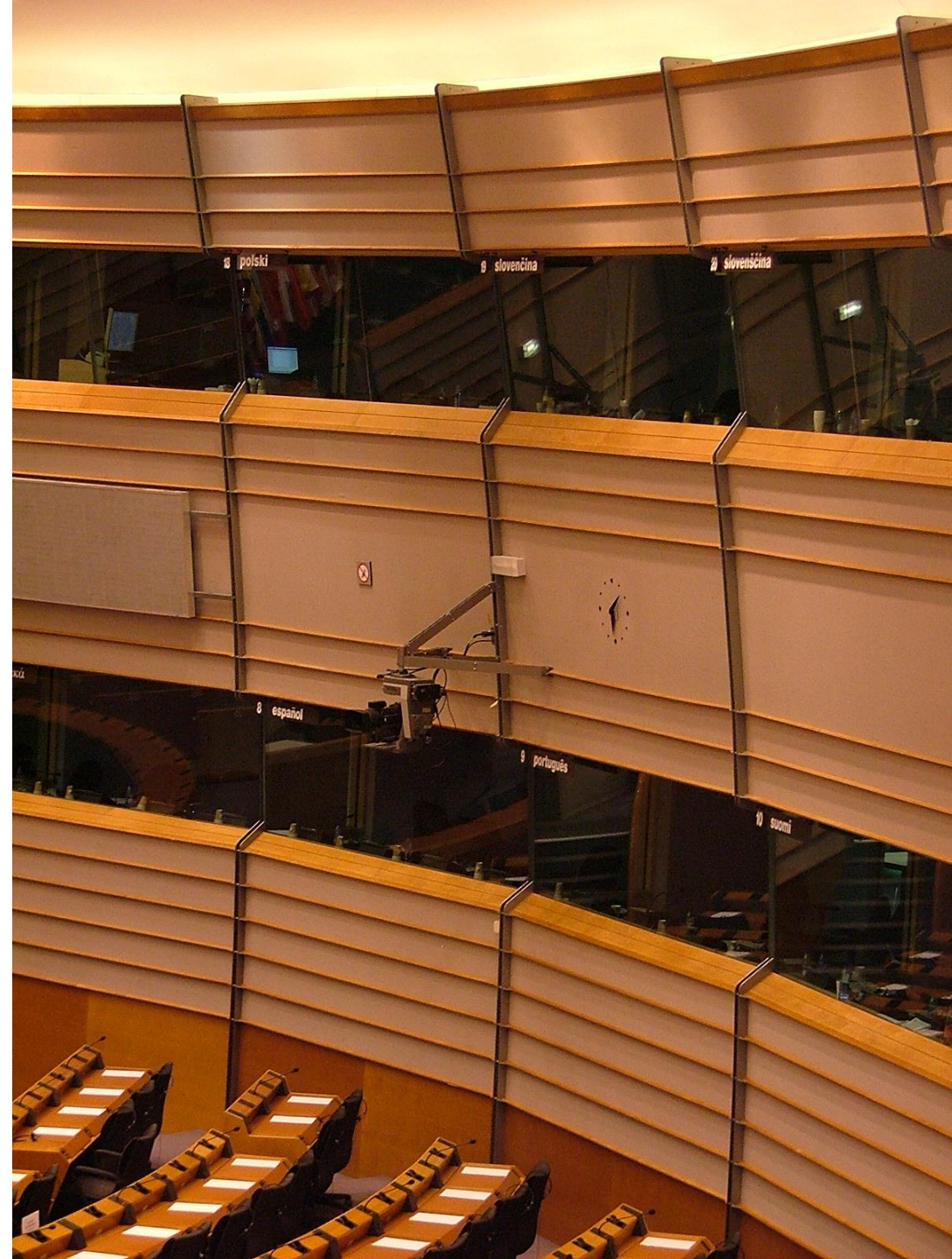
# Breaking down language barriers

Between humans



[photo.catwallpapers.info](http://photo.catwallpapers.info)

Should the interpreter have a persona?





Redmond campus on Sept. 23, 2015. Photo by Brian Smale.

# Why now?

## Confluence of factors:

Steady progress in MT quality over the last few years

- Using vast amounts of data

Technological Leap in Speech Recognition

- Deep Learning (DNNs) – 33+% word error rate (WER) reduction over GMMs (Seide et al 2011)
  - From average of 30% down to 20%, in English
  - Now the improvement is above 42%
- More robust to noise, speaker variation, accents

Skype

- A global platform to put speech translation in the hands of 100s of Millions of users

# Skype Translator: What is it?

1. Conversational Speech Recognition
2. TrueText
3. Machine Translation
4. Skype's Worldwide network
5. A usable human interface

- Oh, I am also very good.

Right now where I'm sitting.  
现在我坐在那里。

Today are.  
今天。

Hello.  
你好。

How are you doing.  
你好。

So I'm doing, well, how are you doing?  
所以我做的, 好, 你怎么样?

Oh, I am also very good.



# How people really speak

## What person thought they said:

Yeah. I guess it was worth it.

→ Ja. Ich denke, es hat sich gelohnt.

→ 是的。我想这是值得的。

## What they actually said:

Yeah, but um, but it was you know, it was, I guess, it was worth it.

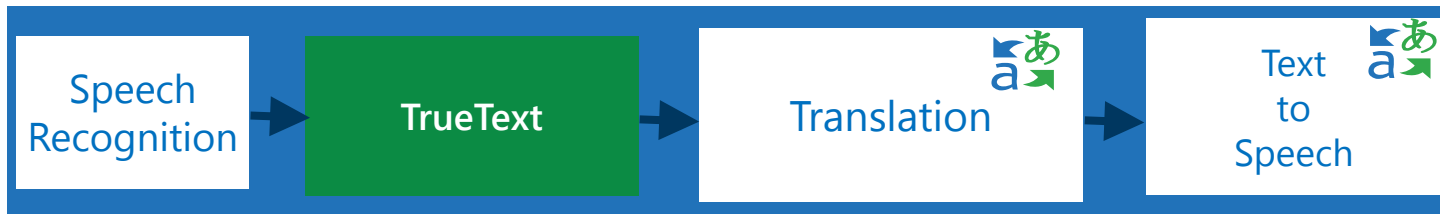
→ Ja, aber ähm, aber es war, weißt du, es war, ich denke, es hat sich gelohnt.

→ 是的但是嗯，但你知道，它是，我猜，它是值得的。

## Disfluency removal

More than just removing "um" and "ah"

**TrueText** is its own translation engine in the same language



**Raw ASR**  
 um no i mean yes but i am i've never done it myself did users before uh I will ask go deep to help me

**Customization and Personalization**  
 um no i mean yes but i am i've never done it myself did users before uh I will ask **gurdeep** to help me

**Lattice Rescoring**  
 um no i mean yes but i am i've never done it myself **did you use yours** before uh I will ask gurdeep to help me

**Disfluency Removal**  
~~no i mean~~ yes but ~~I am~~ i've never done it myself did you use yours before ~~uh~~ I will ask gurdeep to help me

**Segmentation Punctuation and True Casing**  
 Yes.  
 But I've never done it myself.  
 Did you use yours before?  
 I will ask Gurdeep to help me.

Ähm Nein, ich meine ja, aber ich bin, habe ich es nie geschafft mich zuvor Benutzer äh ich werde Fragen gehen tief, um mir zu helfen

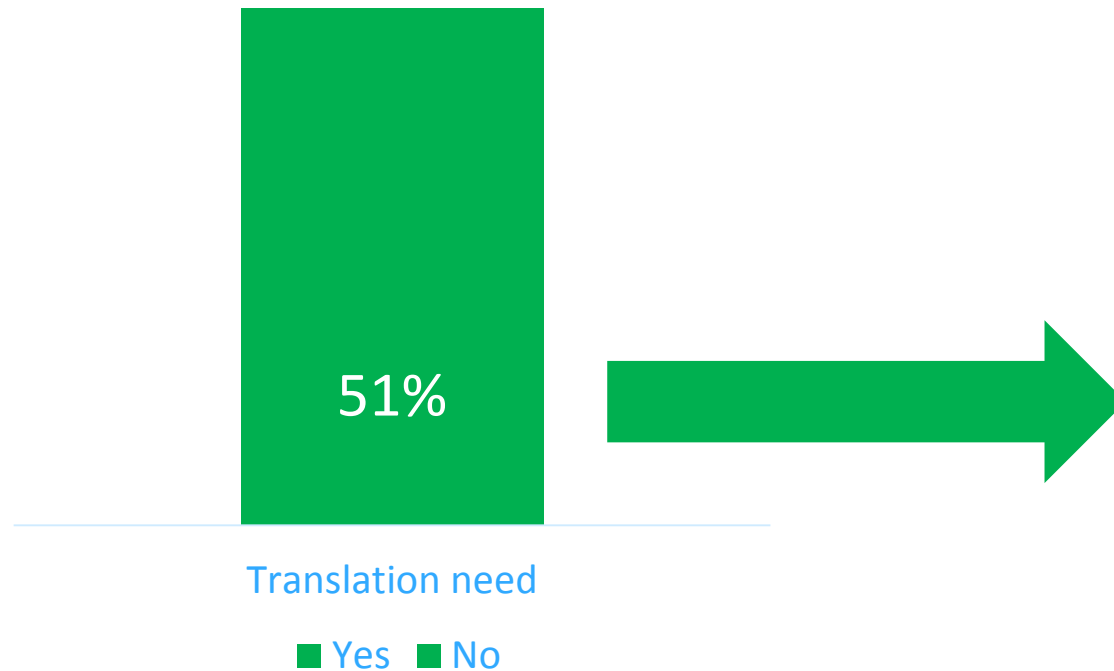
Ja.  
 Aber ich habe es selbst nie getan.  
 Benutzten Sie dir vor?  
 Ich bitte Gurdeep, mir zu helfen.



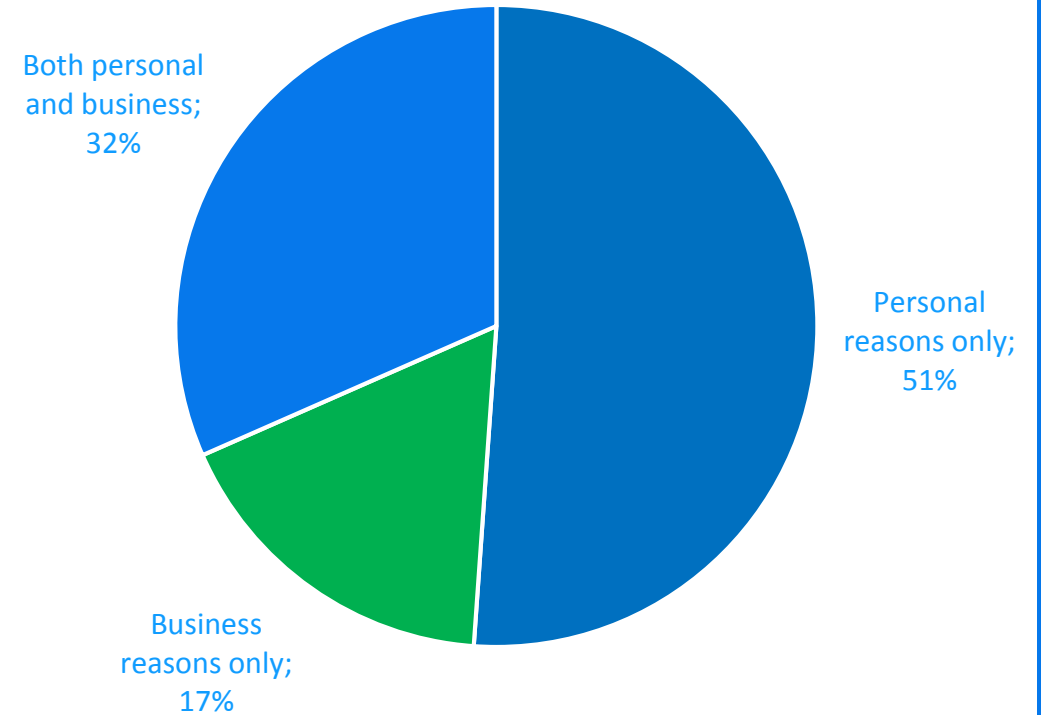
# There is a large perceived need for translation services in the US, across personal and business scenarios

## Current translation requirements

In the last 12 months, have you been in any situation where you wanted or needed to communicate with someone in another language that you are not currently able to speak proficiently or without help?



## Reason for requiring translation

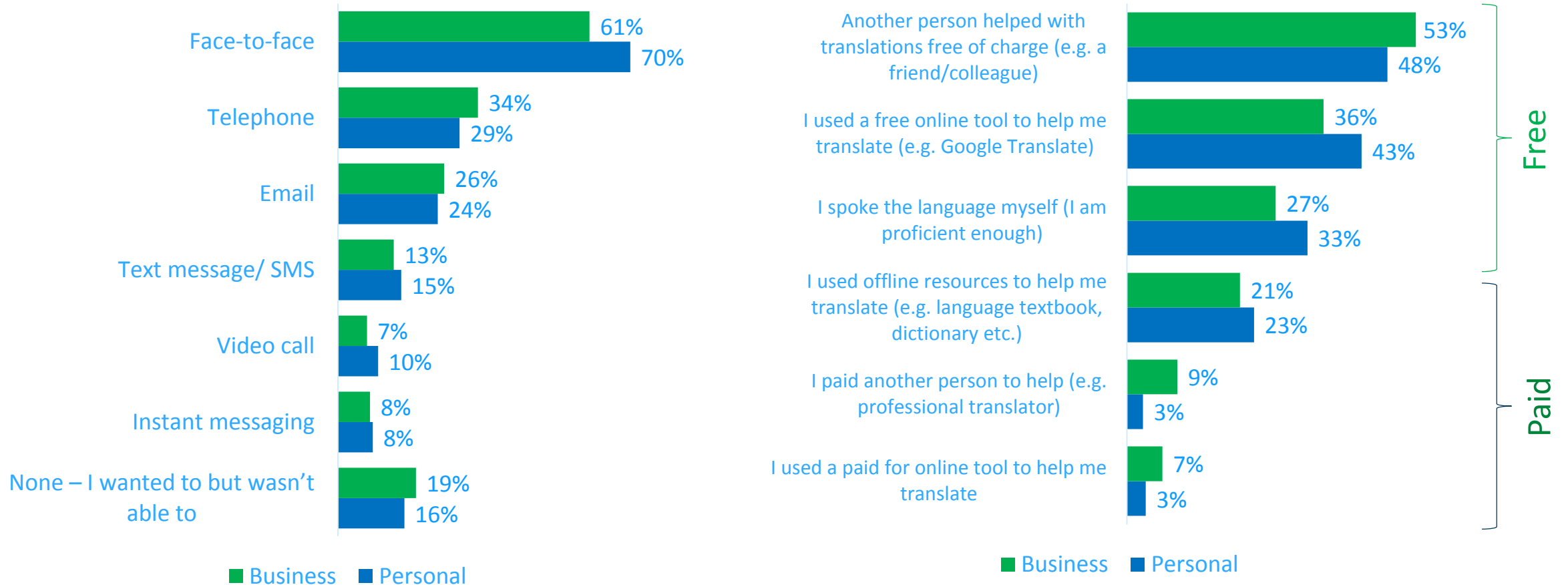


Source: S4: In the last 12 months, have you been in any situation where you wanted or needed to communicate with someone in a another language that you are not currently able to speak proficiently or without help?, S5: And were any of these situations for personal reasons, business reasons or both?

Base: All respondents – Total (1,600), Those who have a need to communicate in a foreign language (1,111)

# The majority of the communication is taking place face-to-face with most using free help for their translation needs

## Current communication habits



Source: A2: How often have you needed or wanted to communicate in another language?, A3: How have you communicated in another language?

Base: All respondents – Those who have a need to communicate in a foreign language (1,111)

## Remote Conversations

1:1 conversation  
2 languages  
2 devices

Skype call with a  
friend

## Brief exchange

1:1 conversation  
1 device

Ordering food in  
Beijing

## Extended social conversations

Many : many  
conversation

Dinner with  
extended family

## Unidirectional Conversations

1: many  
conversation

Classroom or  
Lecture

## Remote Conversations

1:1 conversation  
2 languages  
2 devices

Skype call with a  
friend

## Brief exchange

1:1 conversation  
1 device

Ordering food in  
Beijing

## Extended social conversations

Many : many  
conversation

Dinner with  
extended family

## Unidirectional Conversations

1: many  
conversation

Classroom or  
Lecture



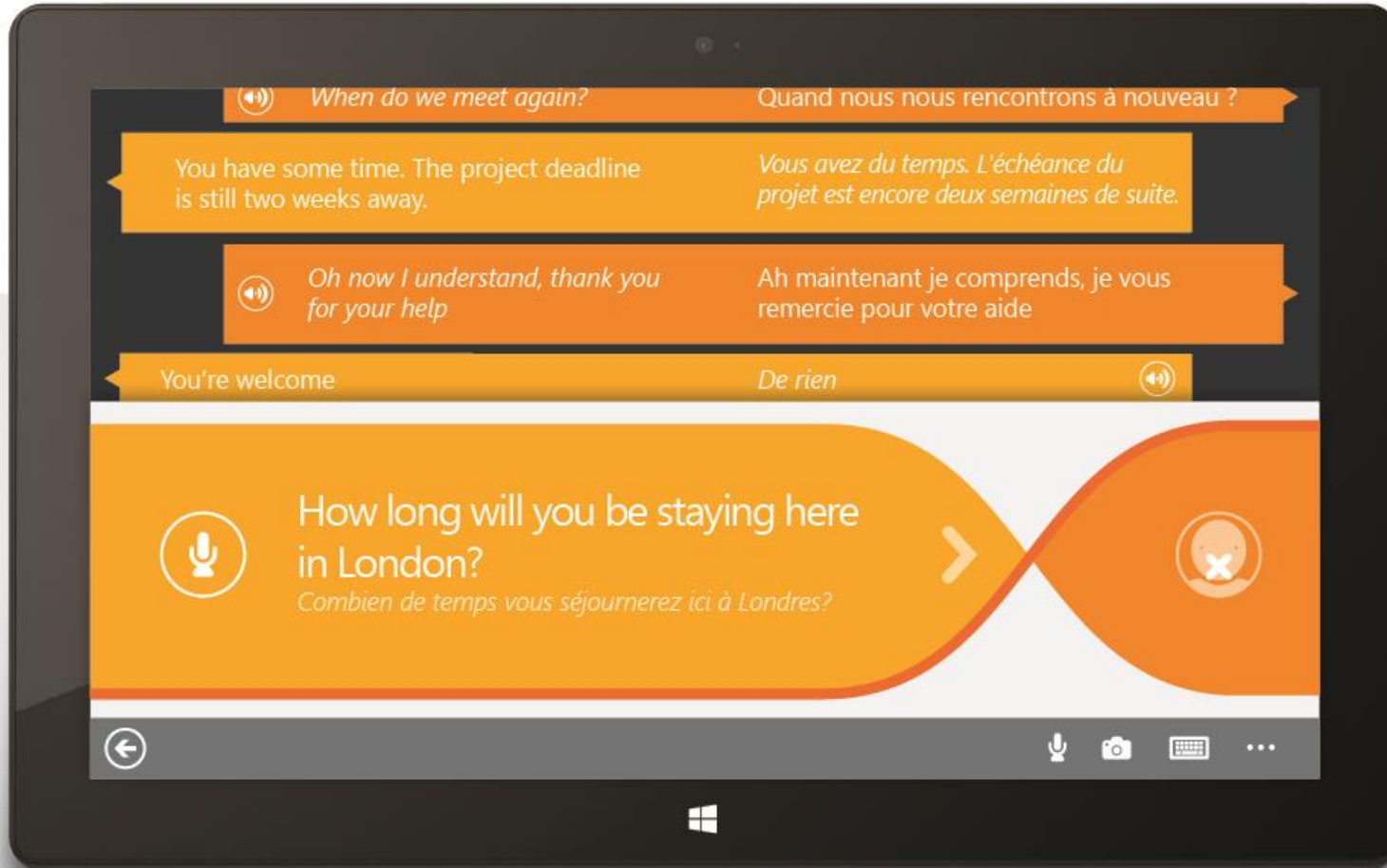




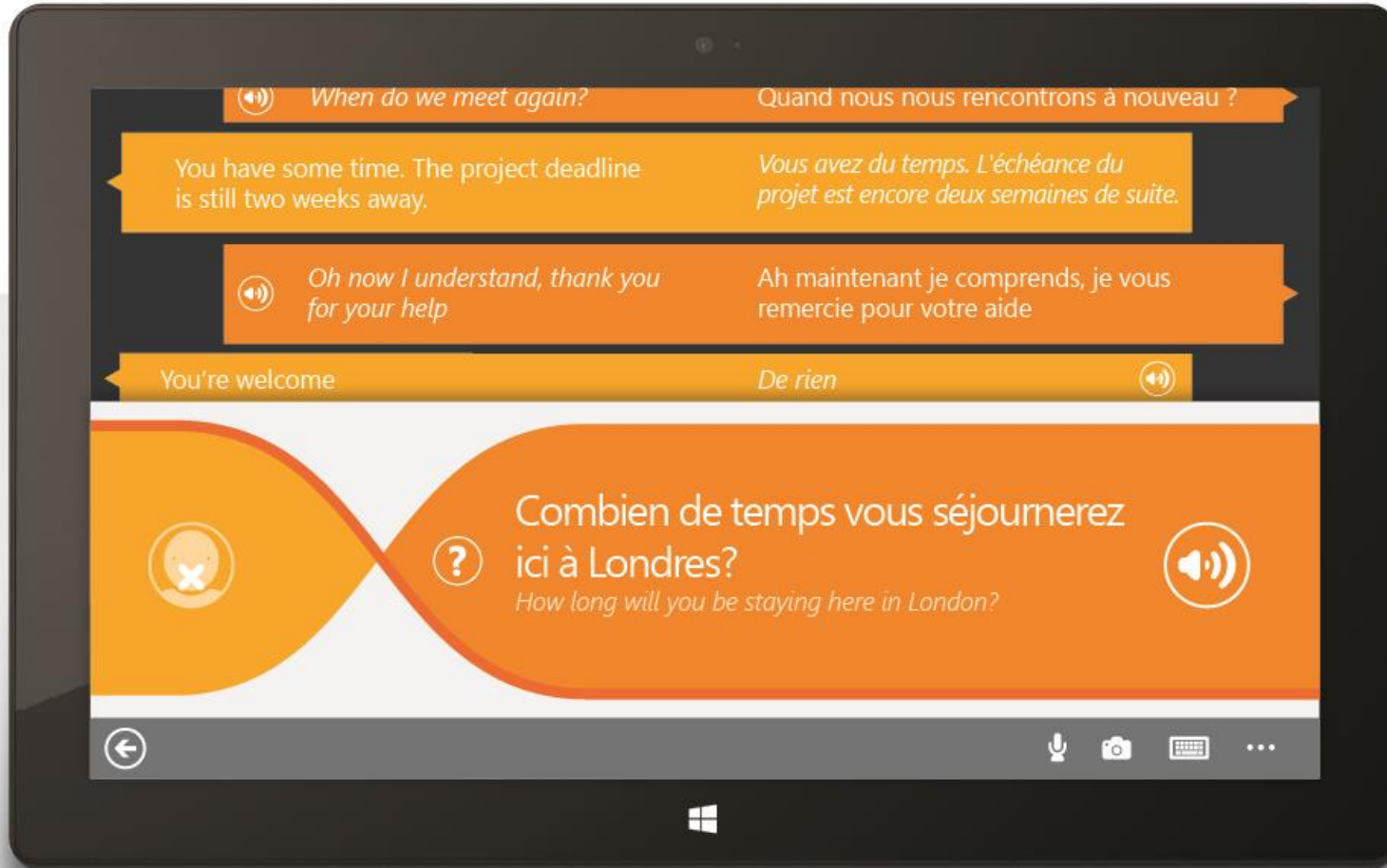




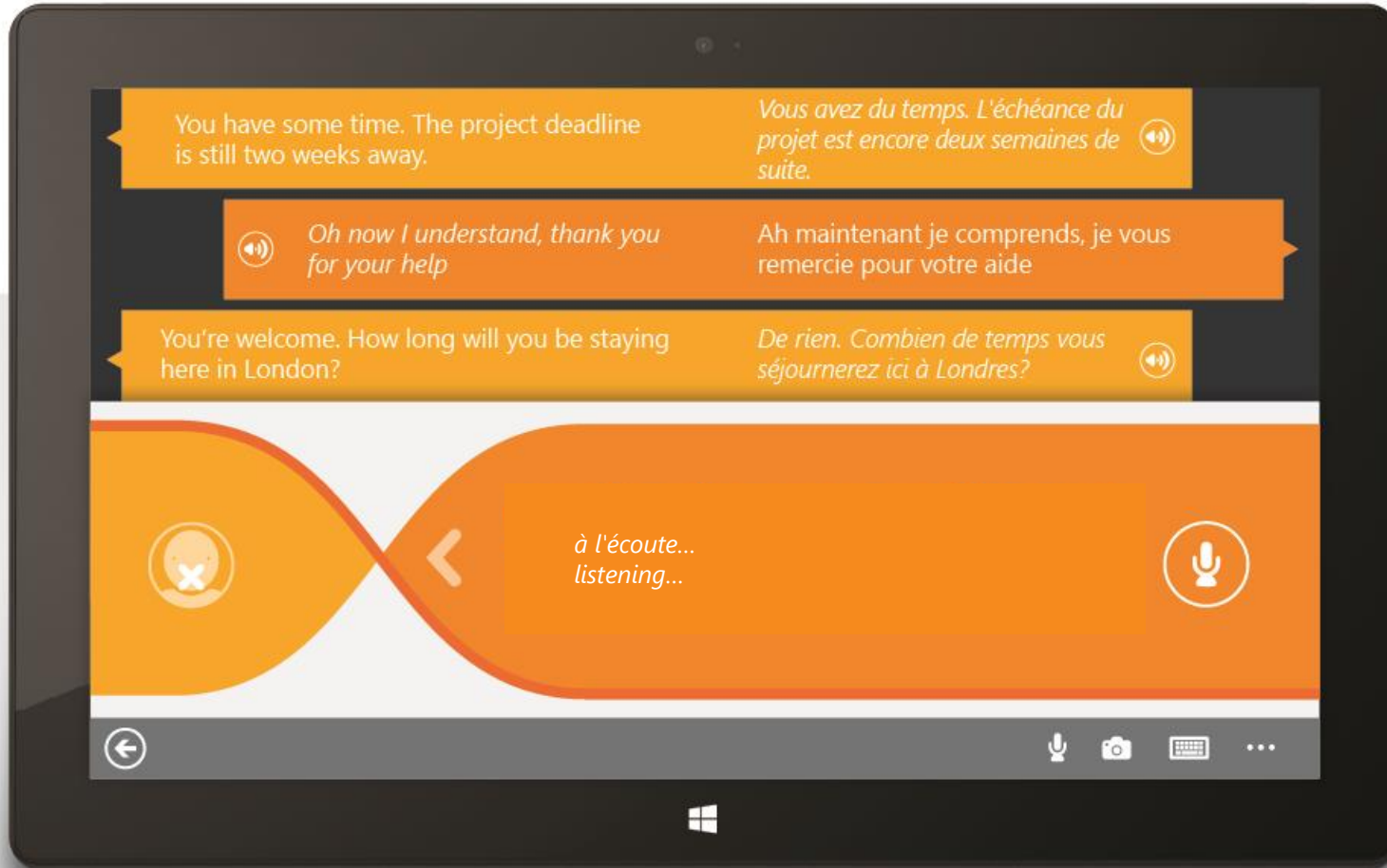




Both active conversation and history include both languages. One will be dominant depending on who is speaking.



Both active conversation and history include both languages. One will be dominant depending on who is speaking.



You have some time. The project deadline is still two weeks away. *Vous avez du temps. L'échéance du projet est encore deux semaines de suite.*

*Oh now I understand, thank you for your help* Ah maintenant je comprends, je vous remercie pour votre aide

You're welcome. How long will you be staying here in London? *De rien. Combien de temps vous séjournerez ici à Londres?*

à l'écoute...  
listening...

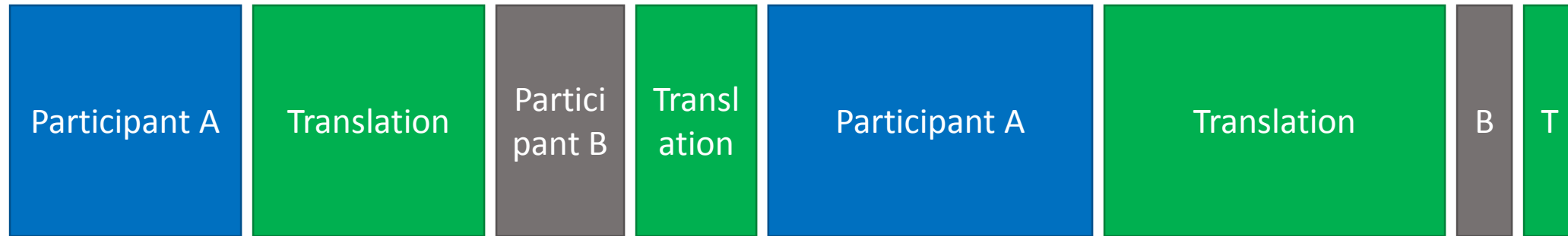
When translated audio finishes playing, app listens for translation.

## User feedback received on that design

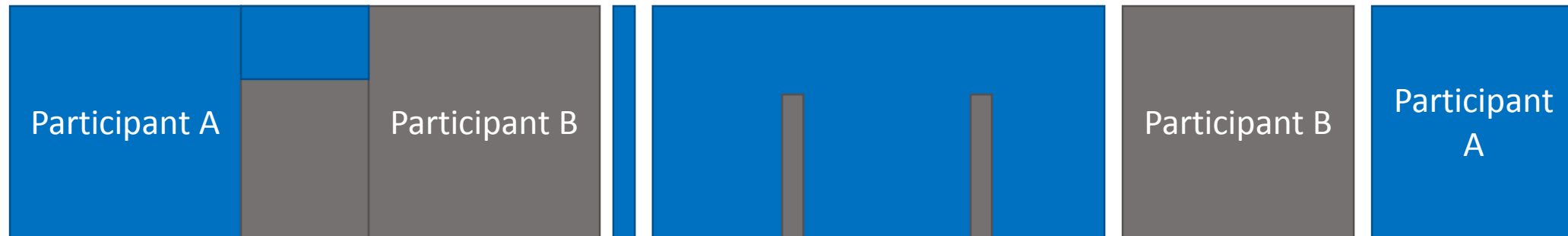
"Skype Translator is also deaf to the rhythms of normal spoken conversation, so you can't be quite sure when its disembodied robot voice is going to break in and start blurting out its translated version. "

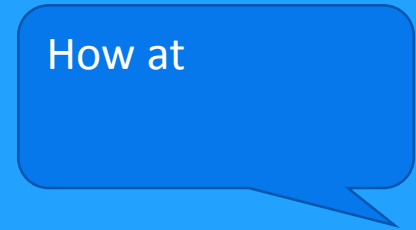
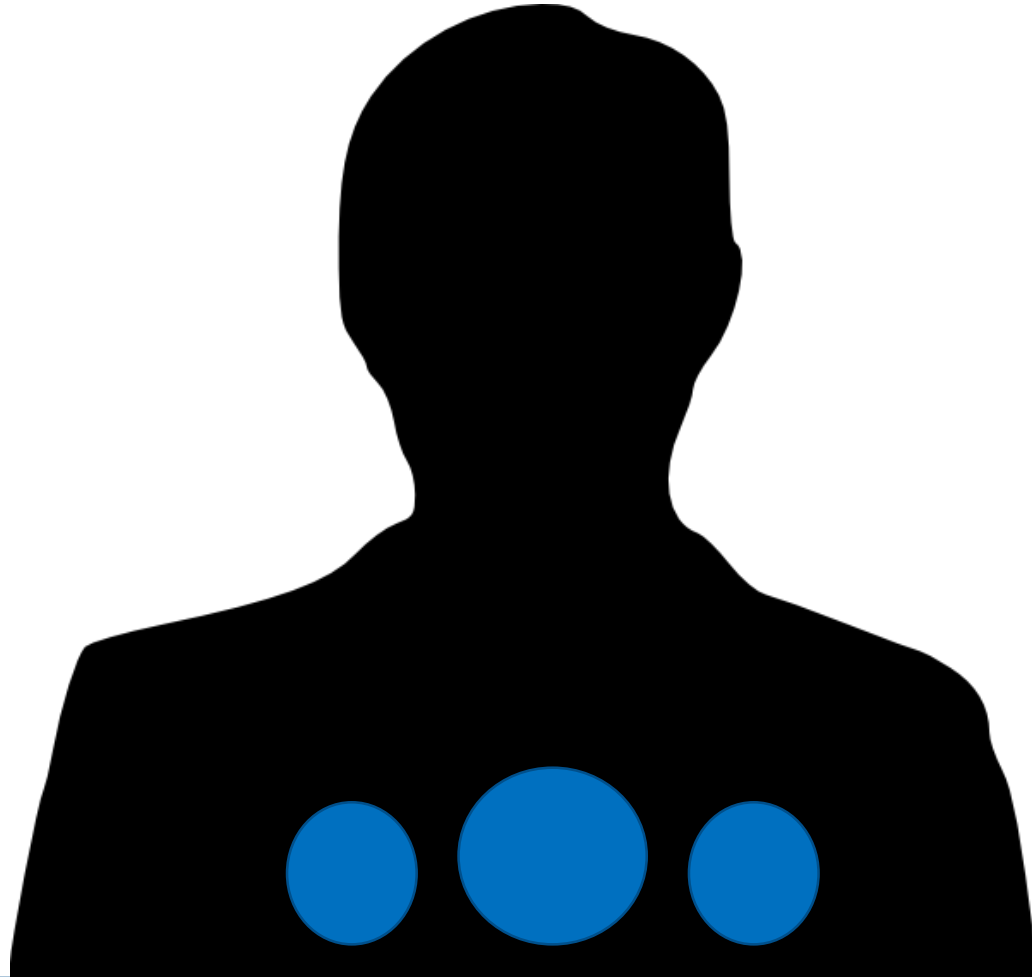
"I know that this is a monumental task and will revolutionize technology... but there isn't a flow in communication ..."

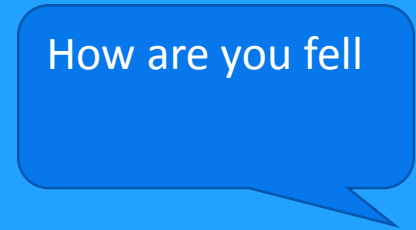
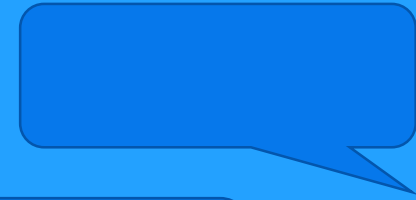
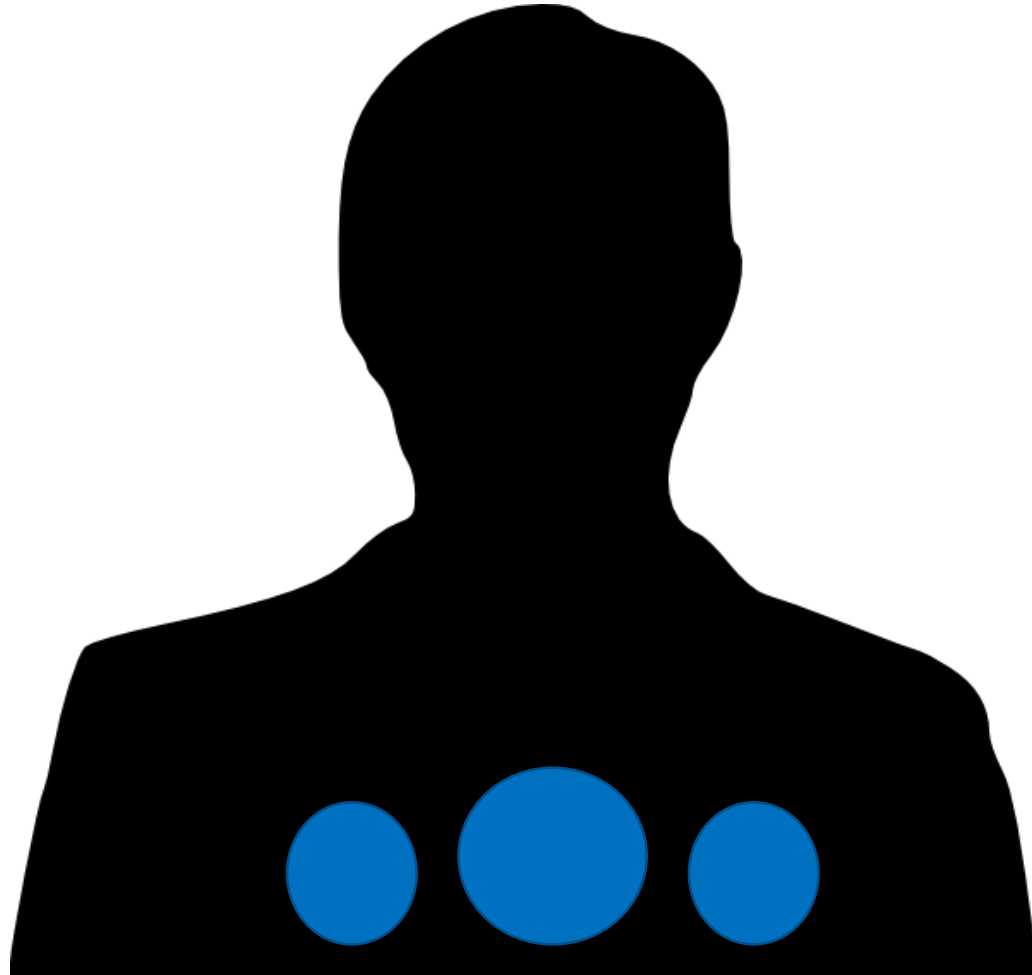
## Skype – defined conversation cadence



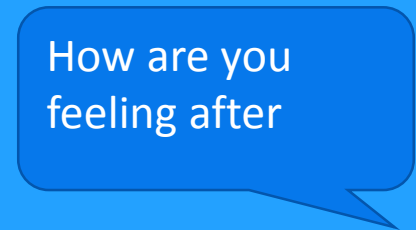
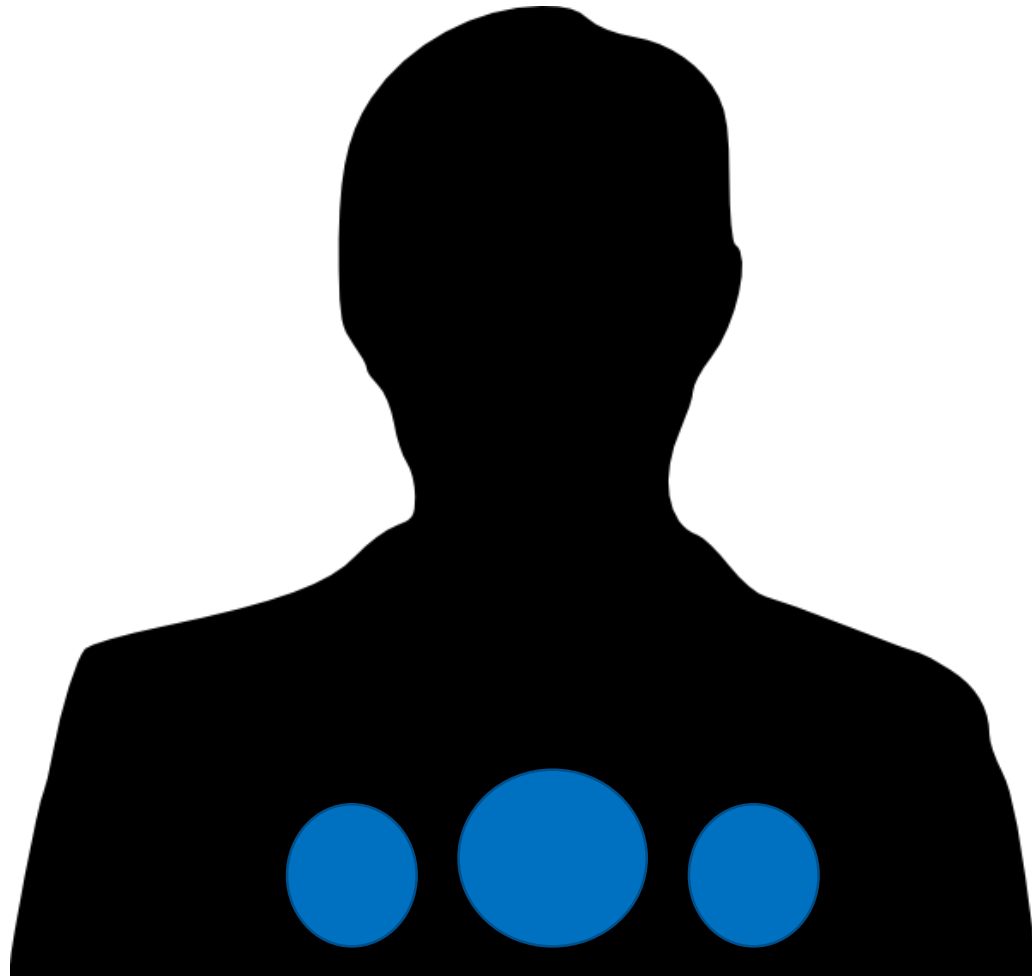
## Natural Human Conversation



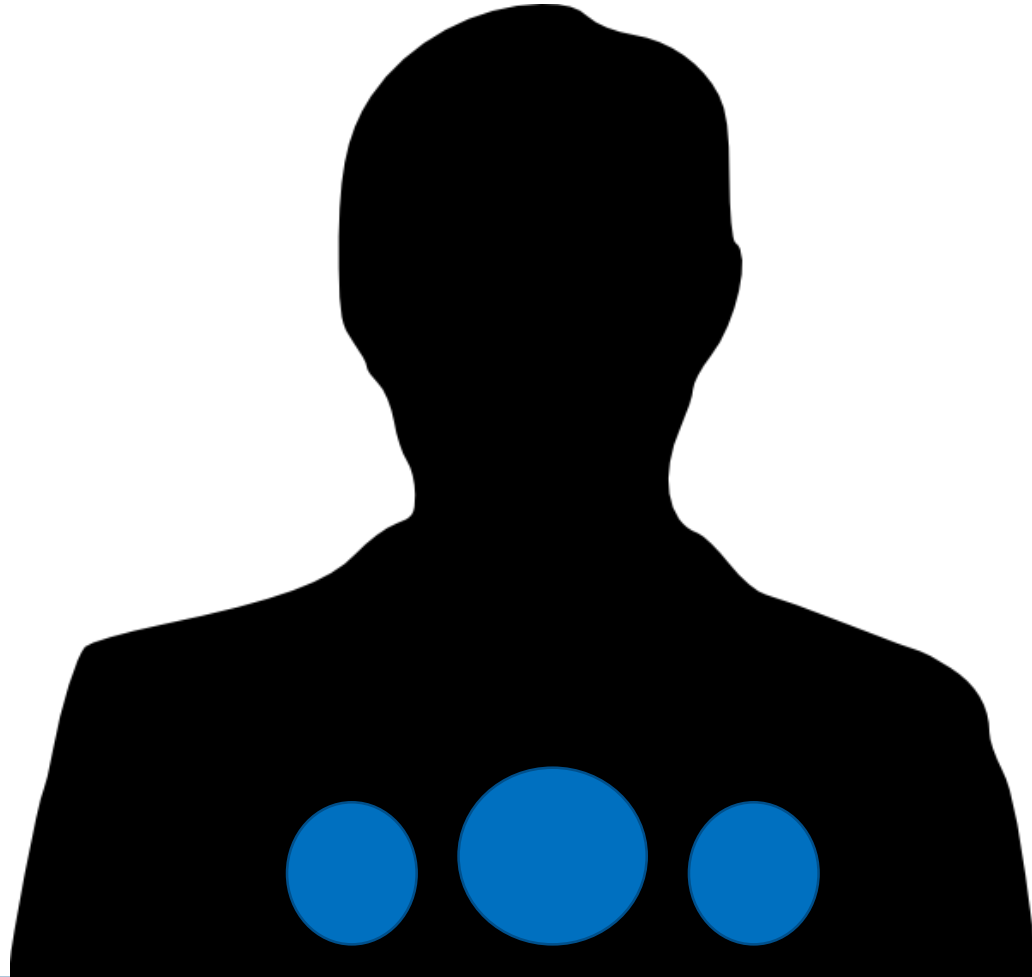








How are you  
feeling after



How are you  
feeling after  
surgery?

Hello!

Hola!

Cómo estás?

How are you?





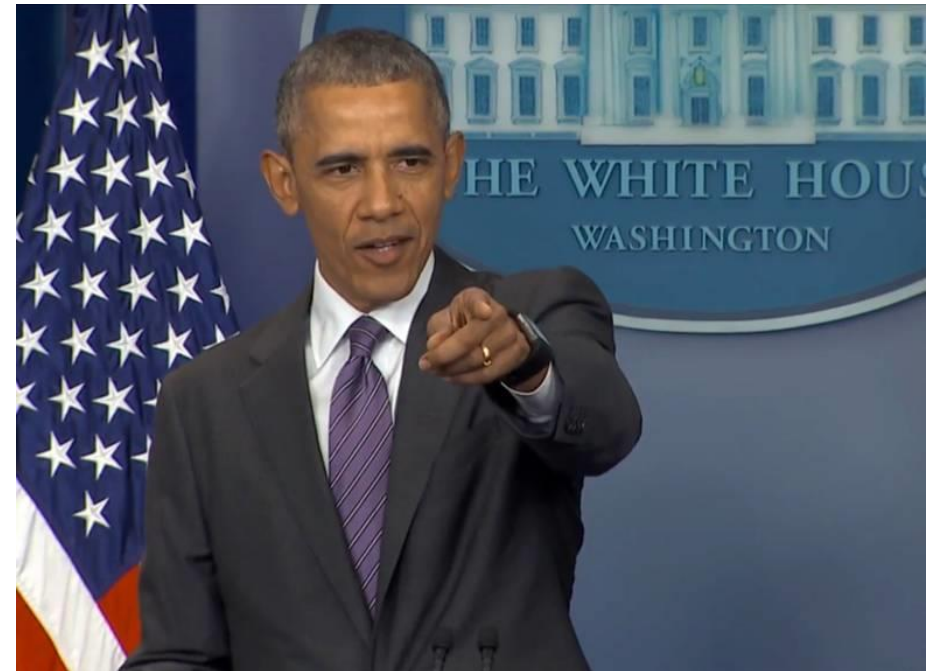
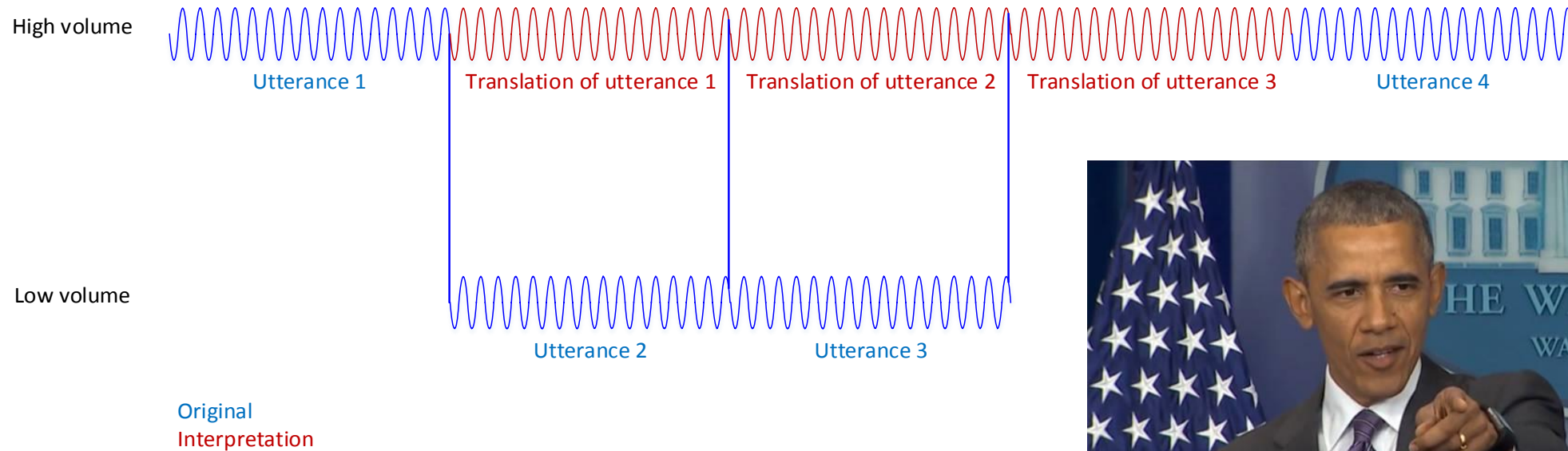
Richard Quest said

Aber der große Unterschied ist die Art und Weise dieser Technologie ist Einsatz sozialer Medien, um den Wortschatz zu erweitern.

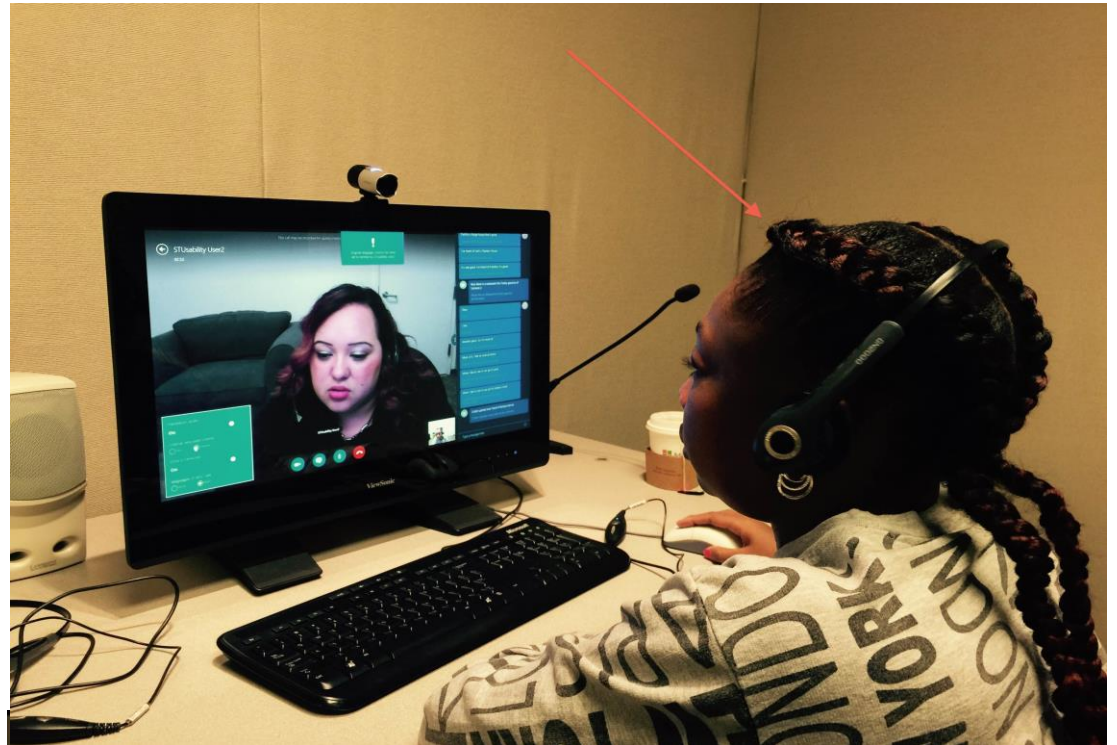
But the big difference is the way this technology is using social media to widen the vocabulary.

This call may be recorded for quality improvement purposes

# Ducking: Varying the volume of the original audio

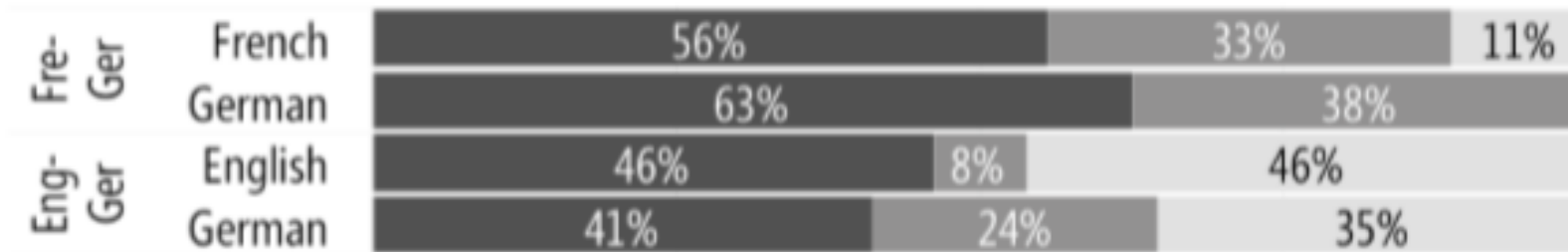


Speaker hears the other person ducked during interpretation.  
Speaker hears his own translation always at low volume.



# Text and Audio User Preference

a. Most Preferred Interface Condition in the Third Round



**Key**

Closed captions and translated audio



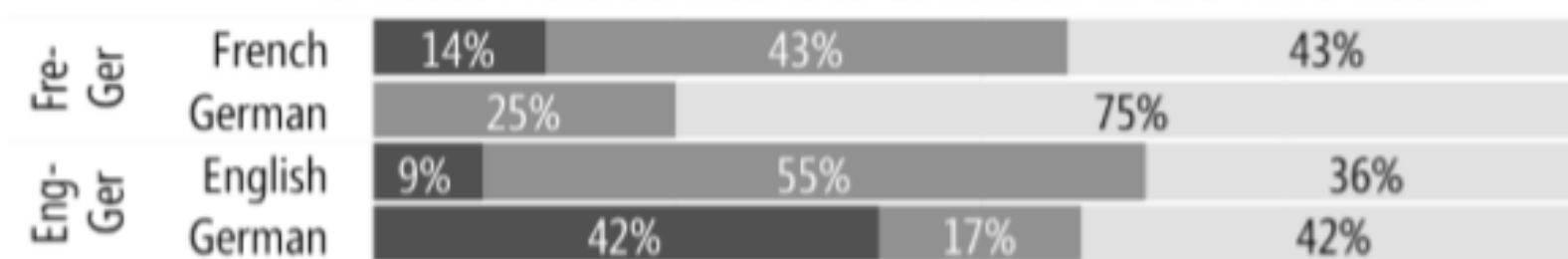
Only closed captions



Only translated audio

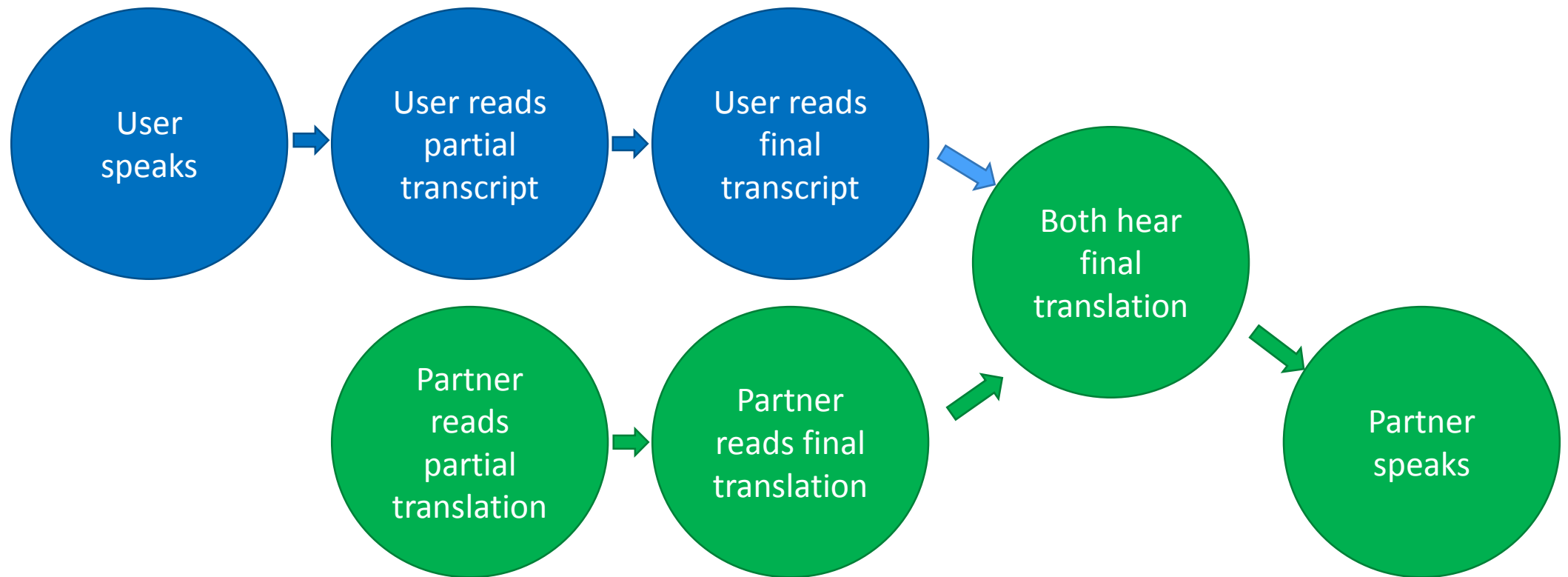


b. Least Preferred Interface Condition in the Third Round



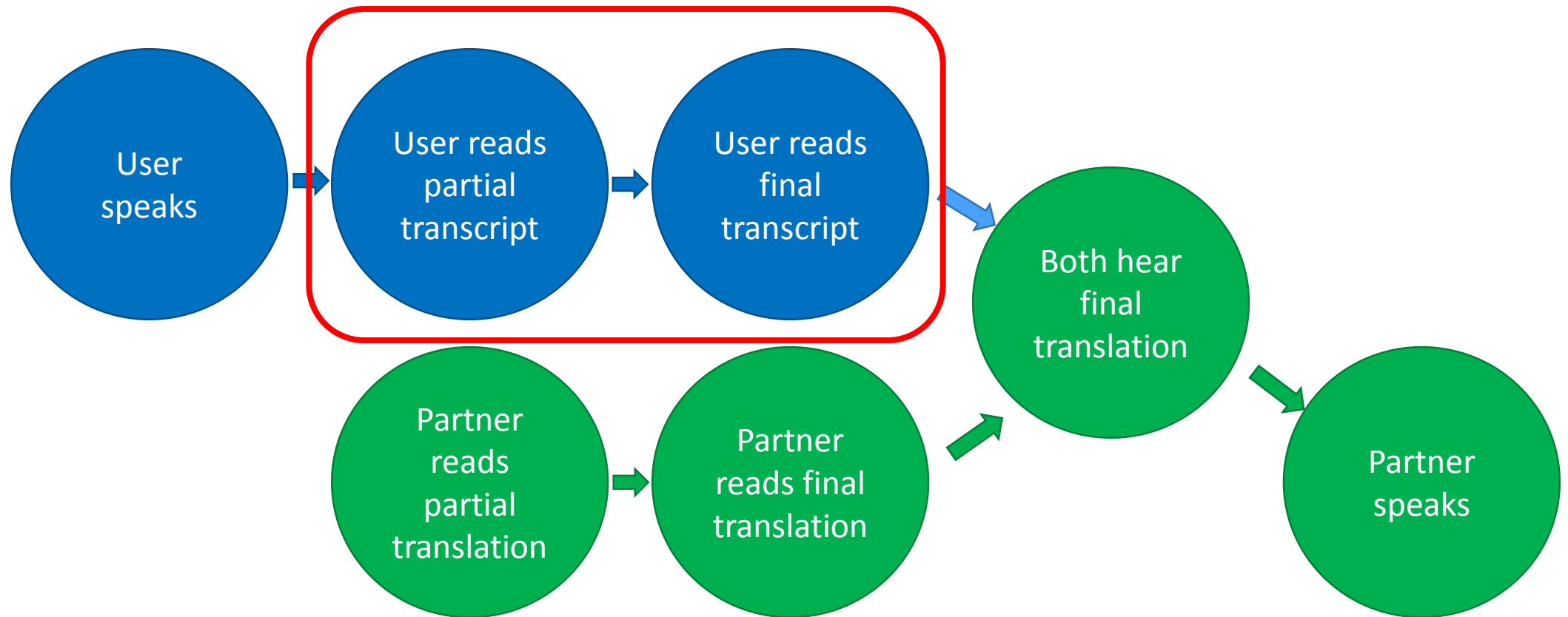
■ CC-TTS   ■ CC   ■ TTS

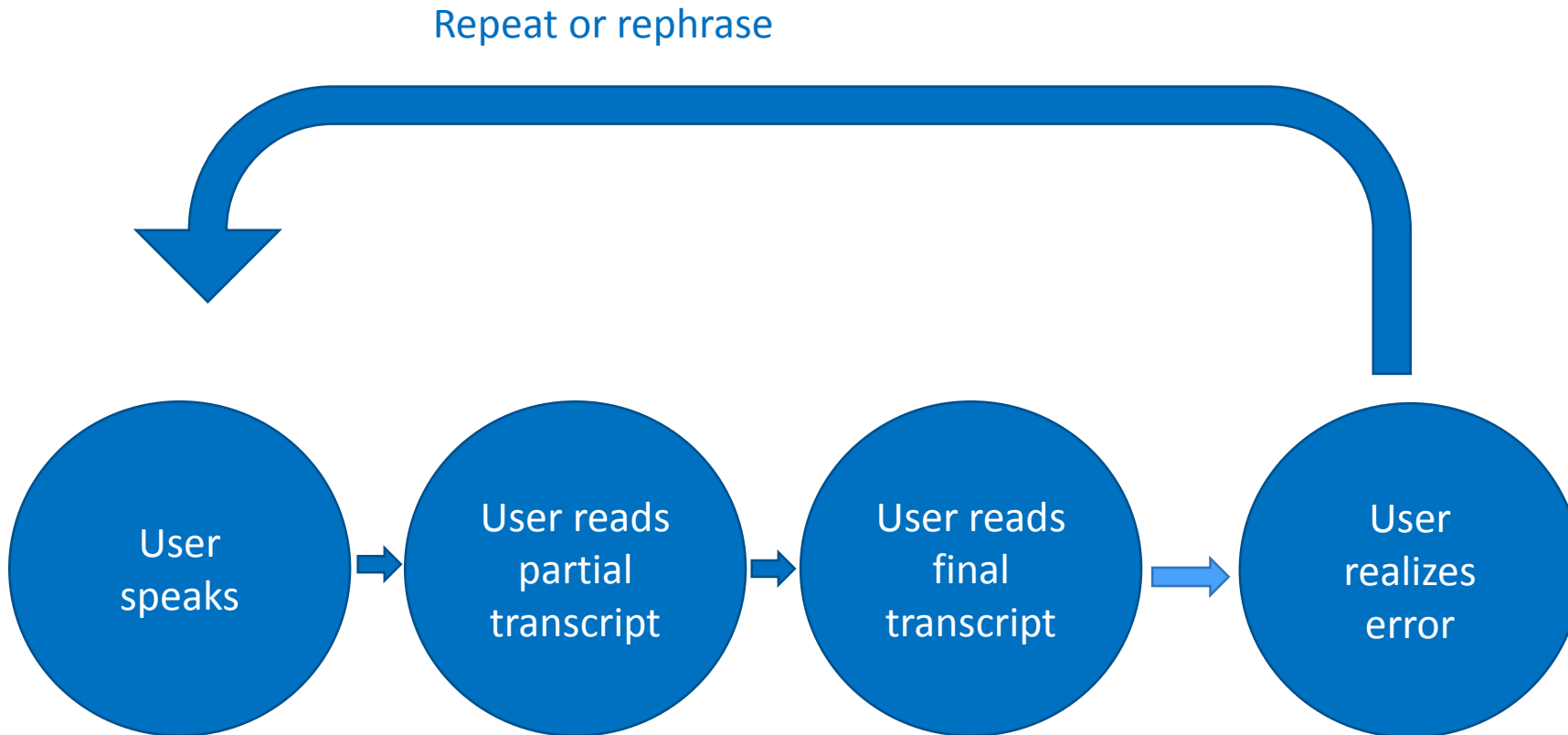
# Typical lifetime of an utterance





# Typical lifetime of an utterance





Repeat  
Rephrase  
Type

[ partial | complete ]  
[ partial | complete ]



skype

Expand



easily edit  
editing....

captains >

captions

capture

chaplains

actions >



original captions  
translated captions

original captions  
translated captions

original captions  
translated captions





This call may be recorded to improve translations.



Sam Smith

02:31

5



TTS

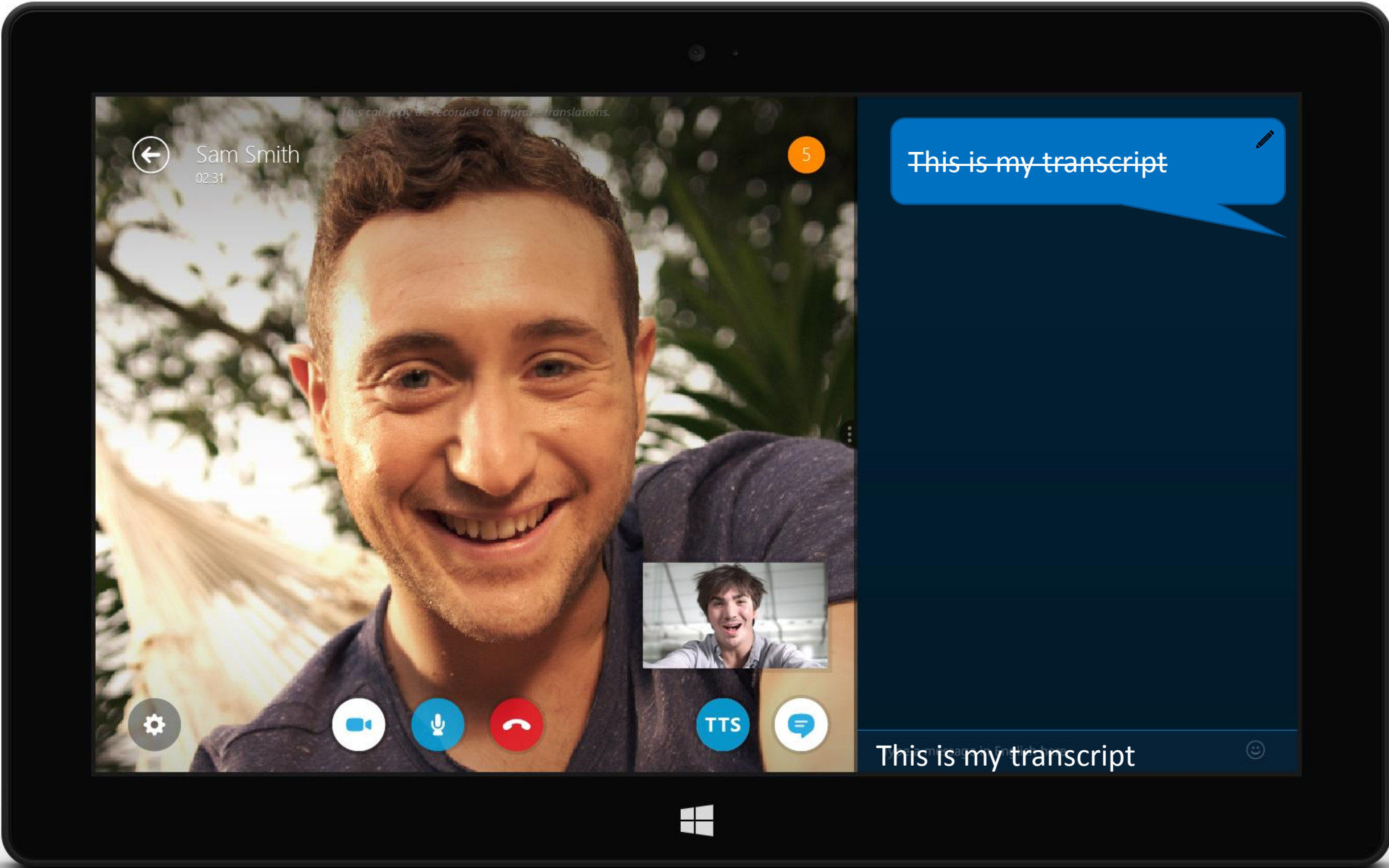


This is my transcript

This is my partner's translated transcript ?

Type a message in English here





*This call may be recorded to improve translations.*



Sam Smith  
02:31

5

This is my transcript



TTS



This is my transcript





*This call may be recorded to improve translations.*



Sam Smith

02:31

5

This is my transcript

This is my partner's translated transcript



Users don't discover that they can type here

Type a message in English here



Where are we going?

# MT quality

- Only one direction: up
- Huge quality differences between languages and content types
- Relatively slow progress
- Neural Networks provide a qualitative jump
  - Especially for languages with very different sentence structures
  - Bringing Chinese<>English up to similar quality as Spanish<>French



# Machine Interpretation

- Starting to become useful for consumers
- Limited business application
- As a help for human interpreters

# Speech Translator on Github

The screenshot shows the Microsoft Speech Translator application window. The title bar reads "Microsoft Speech Translator" with standard window controls. The interface includes several settings sections:

- Account Settings:** Located at the top right, showing the account ID "1958400".
- Microphone/File input:** A dropdown menu set to "Microphone (Realtek High Defini)".
- Speaker:** A dropdown menu set to "Speakers / Headphones (Realtek)".
- From Language:** A dropdown menu set to "English".
- To Language:** A dropdown menu set to "German".
- Voice:** A dropdown menu set to "Hedda (female)".
- Profanity Filter:** A dropdown menu set to "Strict".

Additional controls include:

- Show Subtitles
- # of lines: 2
- Cut input audio during TTS
- Partial Results
- Buttons: Stop, Save logs, Clear logs
- Log options:  Auto-save session log,  Log audio sent,  Log audio received (TTS)

The main text area contains the following text:

This is the application that I am using to translate.

Dies ist die Anwendung, die ich verwende zu übersetzen.

The bottom section displays a log of speech recognition and translation events:

- 07/11/2016 00:18:57.02 Partial recognition 6.5: This is the application.
- 07/11/2016 00:18:57.02 Partial translation 6.5: Dies ist die Anwendung.
- 07/11/2016 00:18:58.09 Partial recognition 6.10: This is the application that I am using.
- 07/11/2016 00:18:58.09 Partial translation 6.10: Dies ist die Anwendung, die ich verwende.
- 07/11/2016 00:18:59.08 Final recognition 6: This is the application that I am using to translate.
- 07/11/2016 00:18:59.08 Final translation 6: Dies ist die Anwendung, die ich verwende zu übersetzen.

The web service API is publicly available.



[www.microsoft.com/translator](http://www.microsoft.com/translator)  
[Translator@Microsoft.com](mailto:Translator@Microsoft.com)

[Chris.Wendt@Microsoft.com](mailto:Chris.Wendt@Microsoft.com)  
Twitter: @Tian500  
WeChat: chriswendt