Contents
Introducing Skype Translator ................................................................. 1
Skype Translator At-a-Glance ................................................................. 3
What to Expect ...................................................................................... 5
Getting Started .................................................................................... 7
Hands-On: Overview and Navigation ................................................. 10
Hands-On: Speech Translation ............................................................. 13
Hands-On: Instant Message Translation ............................................ 17
Summary/Conclusion ........................................................................... 18
Additional Resources .......................................................................... 18
Appendix A: How We Built It .............................................................. 19
Introducing Skype Translator

Did you know that there are roughly 6,500 spoken languages in the world today? There’s even an official Klingon language—not just a few nonsense words invented to make the bumpy-headed, warrior race in Star Trek seem all-that-more-alien, but a complete language created by a noted linguist, with its own vocabulary, grammar, and usage.

If you’ve ever traveled abroad, you’ve likely run into the figurative “language barrier”—a term used to indicate the difficulties faced when people who don’t speak the same language attempt to communicate. Or maybe you’ve encountered it within your extended family, when attempting to interact with certain relatives or in-laws. Similarly, your job may require communication with people around the globe. Either way, whether at work or play, you’ll probably agree that it’s hard to have a meaningful conversation without a common spoken language—that is, until now.

Meet Skype Translator; it helps overcome the language barrier through automatic, simultaneous translation of speech and instant message (IM) conversations. What’s more, it does this at Web scale, delivering these new capabilities within the already-familiar Skype user experience, in a way that’s as intuitive and seamless as possible. December 15, 2014, marks the first time that a preview of Skype Translator is being made available for public use.

To help you understand why we think this is an exciting and newsworthy event, it’s worth taking a brief look at how we arrived at where we are today. Since its inception in 2003, Skype has been about helping people communicate, heart-to-heart and mind-to-mind. Skype is used by more than 300 million people and businesses every month to make free video and voice calls; send files, video, and instant messages; and share both special occasions and everyday moments with the people who matter most. Skype is available in 32 languages and is one of the fastest growing voice communications offerings worldwide, with users spending as many as 2 billion minutes on Skype in a single day.

International use of Skype today is just as impressive, accounting for a significant portion of global long-distance calling. However, while Skype clearly has been transformational in bridging geographic barriers, the language barriers that so often accompany them have, until now, remained obstacles to effective communication.

Skype Translator addresses this gap by augmenting the Skype for modern Windows app with functionality powered by Microsoft Translator, which incorporates more than a decade of effort by Microsoft Research in the areas of speech recognition, automatic translation, and machine learning. In turn, these capabilities are exposed to the user through carefully considered UI enhancements designed to make the experience of a translated conversation as familiar and intuitive as possible. Besides translated conversations, the Skype Translator preview can still do everything that the latest Skype for modern Windows app can.
While we think that you’ll be impressed by what you see, it’s important to note that this initial preview of Skype Translator is just the next step—albeit a significant one—in a longer journey. As we collect feedback from this initial preview, we’ll use what we learn to improve translation quality, fine-tune the experience, and extend the next preview to a broader user base. We’ll also add new languages as they reach a sufficient quality level for productive conversations.

The remainder of this guide should cover everything you need to review the Skype Translator preview: key takeaways, how we built it, what to expect, instructions for getting started, and a “hands-on” guide for experiencing its key capabilities. We hope you’ll come away with an appreciation for how Skype Translator can help people overcome the language barrier, giving them the confidence to have new conversations and explore new parts of the world—in much the same way that GPS and mapping features in modern smartphones and navigation systems have made traveling to unfamiliar destinations less intimidating and more productive.

You may even come away thinking, “Wow... Maybe that Universal Translator that they used to speak with the Klingons wasn’t such a far-fetched, futuristic gadget after all.”
Skype Translator At-a-Glance

If there’s one thing we’d like you to take away from your experience using Skype Translator, it’s that it’s unique in terms of remote, device-to-device speech translation. That said, here’s some other information that you may find useful in experiencing Skype Translator and formulating your review.

Key Features and Capabilities

We’re assuming that you’re already familiar with the current Skype for modern Windows app—the version that runs under the Windows 8.1 modern UI. The Skype Translator preview is based on Skype for modern Windows, to which it adds the following key functionality:

- Speech translation in English and Spanish, which works for both video calls and audio calls
- Instant message (IM) translation in more than 40 languages
- Aids to enhance the translation experience—including automated, simultaneous generation of a transcript in one or both languages
- New settings to support translation—including the ability to turn translation on or off for each contact, and to set each contact’s primary spoken and written languages

In terms of using Skype Translator, here are a few things worth noting:

- Skype Translator is a superset of functionality in Skype for modern Windows; you can still make untranslated calls as you did before—just turn off the Translation switch for that person.
- Only the person initiating a translated call or IM session needs Skype Translator; the recipient can be using Skype Translator or the Skype for Windows desktop client. Other Skype clients (e.g. Windows Phone and Xbox) may also work for the recipients of a translated call, but are not officially supported for this release.

Skype Translator Preview Program

On Monday, December 15, 2014, we formally announced the first phase of the Skype Translator preview program. The purpose of the program is twofold:

- Begin introducing people to Skype Translator and encourage them to sign up for the preview.
- Collect usage data needed to further improve quality and optimize the user experience—the more people use the system over time, the more accurate it will become as we use this new data to improve our language models.

Invitations to participate in this initial phase of the preview program will be offered to selected users who have signed up at www.skype.com/translator.
• To be invited at this time, users must meet the criteria for phase one: an interest in English and Spanish, and a system running Windows 8.1.

• Users invited to participate in the preview will be sent an email message, which will associate the email used to sign up for the preview with the Microsoft Account used to sign in to Skype. The email message also will contain a deep link directly to the Skype Translator preview app in the Windows Store.

• Not all users who have signed up for the preview and meet the initial criteria will receive invitations right away; we may need to limit participation based on system capacity.

The Technology

Skype Translator has been made possible by the combined efforts of Microsoft research, design, and engineering teams around the globe—including more than a decade of work in the areas of natural language processing, speech recognition, machine learning, and automatic translation.

• Core translation functionality for Skype Translator is powered by Microsoft Translator.

• The speech recognition functionality employed by Skype Translator is based on Deep Neural Networks (DNNs), the use of which for speech recognition was pioneered by Microsoft Research. This enables “self-taught” language models that continually improve in accuracy over time, without explicit programming, as new data (audio examples and corresponding, verified transcriptions to text) are made available to the system.

• Translation from one language to another is based on syntactically informed statistical machine translation.

A deeper look at the research, design, and engineering efforts that went into building Skype Translator is provided in Appendix A: How We Built It.

Vision and Future Plans

Initial release of the Skype Translator preview is an exciting milestone for Microsoft. That said, we realize it’s just the start of a journey to transform the way people communicate around the globe.

• Today we’re focused on phase one of the preview; we want to make sure that we’re delivering a great speech translation experience for consumer use in English and Spanish.

• We’ll add support for new languages in the coming months as we reach target quality levels—and expand the preview to include people who have indicated an interest in those languages.

• Our long-term goal is to translate as many languages as possible on as many platforms as possible, and to deliver the best possible experience on each of those platforms for our more than 300 million connected Skype users.
What to Expect

As we've stated, our goal for Skype Translator is to support a new experience—simultaneous, automated translation—within the familiar Skype context. Here are a few things to expect, as well as a few other things that you may find worth knowing.

Test Environment

The accuracy of speech translation can be negatively affected by ambient noise. To obtain the best results, we recommend that you:

- Use Skype Translator in a relatively quiet place, without background noise.
- Use a headset with a good microphone; the one built into your laptop won’t work well. A microphone close to your mouth will pick up nuances in your speech much better, leading to better speech recognition. USB headsets that are Skype or Lync certified are safe bets.

Starting a Translated Call

When you start a translated call, Skype Translator needs to get everything setup for simultaneous translation. Here’s what to expect:

- To hold a translated call, you’ll need to initiate it from Skype Translator. The person on the other end can use Skype for Windows desktop, but it can’t be used to initiate a translated call.
- The back-end process of setting up a translated call can take 30-40 seconds after you initiate it. You’ll need to be patient as your own “virtual translator” is brought into the loop.
- Translation only works Skype Translator-to-Skype Translator, or Skype Translator-to-Skype. It won’t work with SkypeOut, which is the ability to call landline or mobile numbers using Skype.
- Translation only works for one-to-one (two-party) calls, group calls are not supported.

Speaking During the Call

Translated voice calls—especially with speech synthesis turned on—are like working with a human translator; after one party speaks, you’ll need to wait for the translator to generate the spoken translation. Here are some tips for achieving optimal results:

- Speak clearly, with good articulation and at a relatively normal speed; there’s no need to speak very slowly, and you should avoid speaking very fast.
- Avoid using very short sentences. Longer sentences—in the range of 7 to 15 words—provide more context for the system to examine, typically resulting in higher quality for both speech recognition and translation.
• Skype Translator will do its best to wait until you’re done speaking, but any long pauses may cause it to think that you’re finished speaking and ready for it to be translated. Similarly, try to remember to wait for the translation to finish—and be delivered via speech synthesis if you left that setting on—before speaking.

• If you find the delay introduced by listening to the speech synthesis frustrating, try turning off speech synthesis and just following along via the transcript. You can do this during a call from the In-Call Settings—just tap the screen and then tap the Settings button that appears in the bottom-left corner of the app.

• You’ll also find a switch for lowering the original, untranslated voice (technically called ducking) within the In-Call Settings; this decreases the voice of the person on the other end of the call—for example, if you’re speaking English to someone who is speaking Spanish, it reduces the Spanish speaker’s voice to help you focus on the English translation.

Translated IM conversations should be virtually instantaneous. As with speech translation, sentences with seven words or more will often produce more accurate results.

Data Collection
Training Skype Translator to better understand and translate what you say is a continuous machine-learning process; the more people use Skype Translator, the better it will get over time. For this reason, during the preview, we’ll be collecting usage data and recording conversations with user consent. Transcripts will be analyzed for accuracy, corrected if needed, and fed back into the massive corpus of training data so that the system can learn to not make that same mistake in the future.

Other Considerations
Finally, here are a few other things worth knowing:

• Skype Translator may not translate some proper names. For example, it will translate “Spain” to “España” and “United States” to “Estados Unidos,” but it won’t recognize names of places such as “Woodinville” or “Buxtehude”.

• Skype Translator does recognize the names in your address book. If you want to talk about someone and the name is not recognized, try adding that person to your Skype contacts.

• When using Skype Translator, a translated call will look like a regular two-way call, in that you’ll see only the two participants. If the party on the other end is using the regular Skype app or client, that person will see a third “virtual participant” in the call, with an alias such as msmt6536ida. This is the Microsoft Translator service, which actually sets up the call and then invites the person on the other end to join.
Getting Started

Now that we’ve provided an overview of Skype Translator, it’s time to get set up to try it for yourself. The following breaks the process down into two parts: assembling what you’ll need for an optimal experience; and obtaining and installing the Skype Translator preview.

What You’ll Need

To experience Skype Translator, you’ll need the following things:

PC or Tablet Running Windows 8.1
To install and run the Skype Translator preview, you’ll need an x86 or x64 system running Windows 8.1, with the latest updates installed. It also works with the latest Windows 10 technical preview. However, it won’t work with Windows RT.

Headset with Microphone
As we stated in the previous section, speech recognition works a lot better with a headset that puts the microphone close to your mouth; this improves the signal-to-noise ratio and helps minimize room reverberations and other audio artifacts.

Skype Account
We recommend that you try out Skype Translator using your existing, regular Skype account. (If you’d prefer to use a different account or create a new one for this purpose, that should work too. However, you’ll also need to create a new account on your PC because the Skype Translator app inherits the user name from the Microsoft account that you use to sign into Windows 8.1.)

Someone to Communicate With
To experience speech translation using Skype Translator, you’ll need someone who speaks the other language. For example, if you speak English, you’ll want to use it with someone who speaks Spanish—or vice versa. (Remember that only the person initiating the call needs Skype Translator.)

If needed, we can provide Spanish-speaking consultants to help you try everything out. Just send a quick email to skypeinspanish@microsoft.com and we’ll get back to you to get this scheduled.

Initial Setup

After you’ve assembled what you need, all that remains is to obtain and install the preview.

Register for the Skype Translator Preview
For most people, participating in the Skype Translator preview will begin when they register at www.skype.com/translator. As a reviewer, you won’t actually need to do this; we already have your email address to send you an invitation.
Wait for an Invitation via Email
To best experience the Skype Translator preview, you’ll want to uninstall Skype for modern Windows. However, don’t do this just yet. Since you probably won’t want to go without Skype for any length of time, it’s best to wait for the email inviting you to try Skype Translator before uninstalling Skype for modern Windows.

Associate Your Skype ID With Your Microsoft Account
If you’re not already using Skype for modern Windows, you may be signing into Skype using your Skype ID. Skype for modern Windows and Skype Translator only support signing in using a Microsoft account.

Before moving forward with Skype Translator, open Skype for modern Windows—you can find it in the Windows Store, if needed. If your Skype ID is not already associated with your Microsoft account, you’ll have the opportunity to associate them at this time. In the future, with most Skype clients, you’ll be able to sign in using either ID.

Uninstall Skype for Modern Windows
After you’ve received the invitation to install the Skype Translator preview, it’s time to uninstall Skype for modern Windows. To do so, just tap-and-hold the Skype tile on your Start screen or in your Apps, lift your finger, then tap Uninstall. (Remember that Skype Translator provides all the functionality of Skype for modern Windows, so you’ll still be able to make untranslated calls, just like you did before.)

Use “Get Started” Link in Email Invitation to Install Skype Translator Preview
When you receive the invitation to install the preview, click the button within it that says “Get started,” which points to a hidden install point within the Windows Store. If you’re not already logged-in with your Microsoft ID, you’ll be asked to do so before you’re taken to the Windows Store.

On the Skype Translator preview page within the store, tap Install. After the app is installed, you’ll be all set to try it out. (The process of running Skype Translator for the first time is covered in the next section of this guide.)
Tutorials

We’ve produced some videos and help content that you can use to familiarize yourself with Skype Translator. You’ll also find links to them within the app.

Taking Screenshots

If you want to take screenshots, this article provides a few easy ways to do so, without any third-party tools. If you’re using a Surface 3 Pro, you can simply press the volume-down and Windows buttons at the same time, until you see the screen flash briefly.
Hands-On: Overview and Navigation

The following assumes that you already have a basic knowledge of how to use Skype for modern Windows. If not, you can find help and support on using Skype here.

Initial Launch

Tap the Skype Translator preview tile on your Start screen to launch the app. When you do this for the first time, you’ll be asked a series of one-time questions:

- Can Skype Translator preview use your webcam and microphone? → Choose Allow.
- Let Skype Translator run in the background? → Choose Allow.
- Next, you’ll be presented with the terms of use for the preview. → Choose I Agree.
- Choose your language → Select your default spoken language, English or Spanish.
- Choose a voice to represent you → Tap the Jane and Bob icons to hear your options. When you’ve found the one you want, tap Continue.

At this point, you should be looking at the Skype Translator Home screen.

- Note the blue pane that says “Translate calls and IMs.” The links within this pane will take you to the videos and help content mentioned in the previous section.
- As you first use the preview, you’ll see a few help bubbles; each should appear only once.
Adjust Initial Settings

Skype Translator will pick up your existing Skype settings, which are stored in the cloud. For everything to work correctly, you’ll need to check a couple of these settings and potentially adjust them. From within Skype Translator, do the following:

- Swipe left from the right edge of the screen (or move your mouse cursor to the lower-right corner), tap **Settings**, then tap **Options**.

- Under Audio, check the Microphone and Speaker settings; they should indicate your headset. If they say “Use default device,” your headset may not be properly connected and configured.

- Under Translated Calls, note the options for **My spoken language** and **Translated voice** settings; these are what you selected upon running Skype Translator for the first time and can be adjusted here.

- Near the bottom of the Options pane, under Privacy, make sure that **Who can call you?** is set to **Everyone**. The person you’re talking with will also need to have this set to Everyone. (Hint: It’s under Tools > Options > Call Settings if that person is using Skype for Windows desktop.)

Add New Contacts

Now is a great time to add any new contacts you’ll want to call while using Skype Translator. To do so, from the Skype Translator Home screen, just start typing someone’s name, email address, or Skype alias; press Enter; and then tap the button to search Skype.

More detailed instructions on adding contacts can be found [here](#).
Configure Translation Settings for a Contact

From the Skype Translator Home or People screens, tap the tile for a contact with whom you’ll want to hold a translated conversation. While viewing that contact’s profile, note the **Translation** switch in the left pane, under his or her photo and status. Skype Translator adds a switch like this for every contact, with the switch set to off by default.

Turn the Translation switch on, then select that contact’s written and spoken languages. A dialog will pop-up to guide you the first time you turn on the Translation switch for any contact.

- For the written language, you can choose from any of the 40 languages supported by Skype Translator for IM translation.

- If the same language you choose is also supported for speech translation (currently English or Spanish), the setting for the contact’s spoken language will default to the same value. If you chose a written language other than English or Spanish for the contact above, you’ll still need to specify English or Spanish for the contact’s spoken language.

You can turn the Translation switch on or off as often as you want, depending on whether you want to hold a translated or untranslated conversation. When you turn it off, Skype Translator will remember the language settings for that contact for when you turn it on again.
Hands-On: Speech Translation

At this point, you should be ready to hold a translated voice conversation. For purposes of this walkthrough, we'll assume that you're speaking English and the contact you're having a conversation with is speaking Spanish.

Select a Contact for the Conversation

Tap a contact to bring up that person’s profile. If you haven’t already done so, turn on the Translation switch for that contact and set his or her written and spoken languages, as described in the previous section of this guide. (Note that the spoken language you choose will be used to generate both the speech translation and the written transcript.)

Starting the Call

Remember to use your headset for a translated call. Tap the video call button or the audio call button to start the call and then wait for Skype Translator to setup and connect you—remember it will take 20 to 40 seconds.

You’ll know the process is complete when the screen changes and you see that person’s image on the left of your screen.

After your call is connected, you’ll hear the system deliver a prerecorded message (in this case, in Spanish) stating that “You’re about to get an automatically translated message. The message may be recorded by Microsoft in order to improve the quality of the automatic translation service.” After this message is delivered, you’re ready to have a conversation.

If the call times-out and says it was declined, make sure that Who can call you? under Settings > Options is set to Everyone, and that the person you’re trying to connect with has that setting set the same.
Have Your Conversation

To start your conversation—just say something in English. Remember to speak clearly, then pause and wait for the system to perform the translation. It may seem a bit awkward at first, but then again, how often have you had your own personal translator?

Notice how quickly what you say in English is translated into Spanish. Your contact on the other end will hear the same thing you do—that is, your voice in English, followed almost immediately by the translation into spoken Spanish via speech synthesis.

Automated, Simultaneous Transcription

As you’re having your translated conversation, note how Skype Translator generates a real-time, automated transcription on the right side of the screen as you speak. You’ll see what you said in English first, followed immediately by that same phrase translated into Spanish. The person you’re talking with see a similar transcription, only with the Spanish translation presented first.

This transcript can be useful in several ways, such as helping you ensure that Skype Translator recognized what you said correctly. It can also serve as an aid to help you learn the other language.
In Call Settings

When you’re in a call, tap the left pane of the screen (where you see the other person’s profile picture or video) and then tap Settings (the gear icon on the far left) to access the In-Call Settings menu. Here, you can adjust the following:

- **Translation audio**—this is where you can turn translated audio (speech synthesis) on or off. When you turn this off, you won’t hear a translation—you’ll just see the transcript and can follow along that way.

- **Original language volume**—this refers to the volume of the person’s voice on the other end of the call. By decreasing the volume of the language you don’t understand, it may help you to focus on the one you do. It works in both directions, so the other person will hear your voice at a lower level, too. (Technically, this capability is called *ducking*.)

- **Show a transcript**—we added the transcript feature for people who might find it helpful. If you don’t need it or just find it distracting, feel free to turn it off.

- **Languages of text I see**—this applies to the simultaneously generated transcript. You can choose to display both languages for each translation, with the one that you speak presented first, or to show only the language that you understand.

Feel free to play around with these in-call settings during a call. We implemented them as options for a reason, realizing that the optimal experience for everyone may not be the same—and may change over time as one becomes more familiar with Skype Translator.
End the Call and Save a Transcript

To end the call, tap the screen like you did to access the In-Call Settings and then tap the **end** button—it looks like a red telephone. Your screen will revert to the profile for that contact, with his/her picture and language settings on the left and a conversation history on the right.

In the conversation history, you’ll see an entry for the call you just completed. You can save a transcript of the call by tapping the **save transcript** text within that entry. To change the location where you want to save the file, after tapping **save transcript**, just tap the down-arrow next to the current location specified—it’s at the top-left of the screen and should initially default to OneDrive.

Below the save transcript text, you’ll be presented with the opportunity to rate your experience on a scale of one to five stars. If you provide a rating of less than four stars, you’ll be presented with the opportunity to provide more detailed feedback. As we’ve said, one of the major goals of our preview is to collect this type of quality information, analyze it, and use it to further improve our machine learning models.
Hands-On: Instant Message Translation

Holding a translated IM conversation is even easier than a translated voice call; just choose a contact, make sure the Translation switch for that person is on, and start typing a message. When you hit enter (or tap send), you’ll see your original message in the right-hand pane, followed by its translation. Your contact on the other end will see something very similar, albeit with the translated message in the language that he/she understands presented first.

Here are a few things worth noting:

- Skype Translator will use the written language settings for the contact (and for you) to determine how to perform the translation.

- IM translation is virtually instantaneous, and you won’t experience the same setup delay as for a translated voice call—making this a great way to get a quick answer to a question or just send a quick note to someone, without the language barrier getting in the way.

- IM translation is supported in more than 40 languages, so feel free to experiment with languages other than English and Spanish. Yep… even Klingon.
Summary/Conclusion

For more than a decade, Skype has been about helping people communicate—heart to-heart and mind-to-mind—across geographic barriers, no matter how large or small. With Skype Translator, we’re taking the next step in this journey by tackling the language barrier, opening up new ways to make meaningful connections and hold meaningful conversations across the entire spectrum of human relationships—including friendships, families, business, education, diplomacy, and more.

In our industry, we often talk about pursuing big, bold dreams, and of how we’re limited only by the power of our imaginations. Skype Translator is one of those endeavors and, as we proceed with the preview program and toward formal release, we look forward to keeping you apprised of our journey to break down yet another barrier to human productivity and connections.

Additional Resources

Following are some additional resources that may be helpful:

- **Skype Big Blog** – including company updates, big product releases, and some of our most popular blog posts.

- **Skype Garage and Updates Blog** – where you can learn about the newest features, products, and client updates for all platforms.
Appendix A: How We Built It

Skype Translator has been made possible by the combined contributions of Microsoft research, design, and engineering teams around the globe. We view each of these three areas as equally essential in delivering a natural and accurate translation experience.

Language and Speech Research

Skype Translator builds on more than a decade of progress by Microsoft Research in the areas of speech recognition, automatic translation, and machine learning. However, research into machine-powered speech translation dates back much further. In brief, here are the major milestones:

- During World War II, success in breaking wartime cryptographic codes led to the belief that similar methods might be successful in translating from one human language to another.
- In 1954, IBM and Georgetown University demonstrated a computerized translation system based on six grammar rules and a 250-word vocabulary.
- In 1966, various U.S. government agencies formed the Automatic Language Processing Advisory Council (ALPAC) to study machine translation.
- In 1975, researchers proposed what has become the basis for modern speech recognition systems, and research began at IBM and Carnegie Mellon University.
- In 1990, IBM researchers published a seminal paper on machine translation; it suggested using statistical methods based on the analysis of large amounts of data over earlier approaches.
- In 1997, Dragon Systems and IBM released the first commercial software for large-vocabulary, continuous speech recognition—the first time such functionality was broadly available.
- In 2007, Microsoft launches Windows Live Translator (later Bing Translator)—made possible by Microsoft Research breakthroughs in syntax-based statistical machine translation.
- In 2011, Microsoft showed how the use of Deep Neural Networks (DNNs) for speech recognition could deliver substantial improvements in accuracy—the largest in decades.
- In 2012, Rick Rashid, at the time head of Microsoft Research, demonstrated the combined potential of these breakthroughs at a conference in China. When Rick spoke in English, the system delivered the translation to Chinese using Rick’s voice—leaving the crowd speechless.
- Satya Nadella and Gurdeep Pall (Microsoft CVP for Skype and Lync) first demonstrated Skype Translator publically in May 2014, at the inaugural Code Conference in Rancho Palos Verdes.

A full timeline of the major research milestones that have contributed to the delivery of Skype Translator can be found here.
User Experience Design

Since its initial release in 2003, Skype has always put the user experience first and foremost. We took a similar approach with Skype Translator, realizing that we had to take into account the unique aspects of an entirely new mode of communication. We carefully considered the many “human” elements of speech translation, leading us to ask ourselves questions such as:

- How can participants in a translated conversation ensure that Skype Translator accurately recognized what they said—let alone translated it correctly?
- What’s the best way for participants in a translated conversation to wait for the translated speech to arrive?
- How do we reduce the confusion that might ensue when a participant understands the original utterance and replies before waiting for the translated speech?

Extensive discussion around these types of questions drove many of our design decisions, leading to features such as the In-Call Settings menu, where you can tailor the experience to your liking. For example, during a translated voice call, you can:

- Choose whether to see a simultaneous, automatically generated transcript—and whether you want it to display only the translation or both the original and translated speech.
- Choose whether you want to hear the translation via speech synthesis, or turn it off and just rely on the transcript.
- Turn ducking on or off. (Ducking lowers the volume for the untranslated speech, so that the translated speech is louder and thus easier to follow.)

We’ll continue to refine the design of Skype Translator and the user experience it provides based on what we learn during the preview.

Engineering

The engineering effort to build Skype Translator required addressing a range of challenges. Perhaps more important, the solutions to those challenges had to be integrated into a cohesive whole—one that, while it may represent quite complex underpinnings, is still capable of delivering a simple and straightforward user experience.

Following is a closer look at that architecture, as illustrated in the below diagram:
Training Data for Machine Learning

Microsoft Translator applies machine learning techniques to train and optimize its automated speech recognition and machine translation models. The training data for these tasks comes from a variety of sources, including translated web pages, videos with captions, and previously translated and transcribed one-on-one conversations.

After the data is prepared and fed into the machine learning system, it builds statistical models of the words in these conversations, along with their context. In turn, these statistical models teach the speech recognition engine how to map the incoming audio stream to text, and they teach the machine translation engine how to map the text to the phrases of the other language.

Automated Speech Recognition

When you say something, Skype Translator looks for something similar in its statistical model and applies the previously learned transformation from audio to text. While speech recognition has been an important topic of research for decades, until recently, its widespread adoption had been stymied by high error rates, sensitivity to variations in the speaker’s voice, ambient noise, and so on. The use of Deep Neural Networks (DNNs) for speech recognition, pioneered by Microsoft Research, resulted in dramatically lower error rates and improved robustness, finally enabling the use of speech recognition in broad contexts such as Skype Translator. At the same time, the dream of global human-to-human communication was major motivating factor and driving force for the Microsoft researchers working on this technology. Over the course of this preview, Microsoft hopes to further improve the ability of its language models to handle a variety of speakers, accents, and conditions.

Speech Correction

Before text output from the speech recognition engine is fed into the machine translation engine, it is corrected and transformed to facilitate translation—a task made necessary due to the surprisingly large difference between how people write a language and how they speak it. This correction
process includes the removal of disfluencies, such as the ‘ahs’, ‘umms’, pauses, and rephrasings that often creep into conversations as people think, correct themselves, and change their minds when speaking—none of which should ultimately appear in the translation or its accompanying transcript. The speech correction process also includes division of the text into sentences, as well as the addition of punctuation and capitalization.

Text-to-Speech

Speech synthesis, the final step in the speech translation process, happens when the speech synthesizer receives a phrase or sentence for conversion to audio. This system component is based on existing, relatively mature technologies.

Skype Translator gives each user the option of sounding like a male or like a female speaker, independent of the gender in that person’s Skype user profile. Speech synthesis also is optional: if conversation participants are at ease following each other’s utterances using the written transcript, they can turn off the translated audio. This facilitates a smoother conversation because most people can read much faster than they can follow-along by listening.

Usability research for Skype Translator has found that participants in a translated conversation who do not understand the other language at all typically like the audio translation, whereas those who have some familiarity with the other language often prefer the faster-paced conversations they can have without it. We left audio translation on by default in the preview because it makes conversations easier when using types of devices where following the text is more difficult than just listening to the audio, such as mobile phones.

Machine Translation

The machine translation part of Skype Translator translates text from one language to another. Skype Translator uses the same technology that powers Bing Translator on the web, which pioneered the combined use of syntax and statistical models specifically trained for a conversational type of language. This was particularly challenging, as the typical training data used to build text translation systems today is optimized for clean, well-formed written language. Microsoft Translator addresses this by combining the broad language knowledge of Bing Translator together with an extensive layer of the words and phrases that are used in spoken conversations.

The translation engine also addresses the fact that spoken languages can often reflect our regional, national, and cultural identities through colloquialisms or slang. Fortunately, Microsoft Translator has become highly adept at colloquial translation due to its years of use across various social media sites, including Facebook and Twitter. Data collected during its preview phase will enable the system to observe and learn from additional levels of casual conversation, bringing it even closer to accurate translation of truly conversational speech, including casual phrases and terminology.

Skype Translator ‘Bots’

Engineering teams used customized ‘bots” to orchestrate the entire experience. These bots are responsible for creating the call, and for sending audio streams to the speech engines in exchange
for translation and transcription. The translator bot acts like a third participant (i.e., your own virtual translator) in the call; it translates what you just said when you've finished talking, and it translates what the person you’re speaking with just said when he or she finishes. Experience has shown that people who are used to working with a human interpreter are quickly at ease with this mode of interaction, whereas others may require some time to get used to it.

More Work to Do

In conclusion, while the initial release of Skype Translator preview represents a major milestone for all of the research, design, and engineering teams across Microsoft that have contributed to it, we also all realize that the preview is simply another step in delivering the best possible translation experience. We'll rely on the feedback and data that preview users share with us to help improve the technology, refine our approaches, and ultimately optimize the unique experience that Skype Translator provides. It's still early days for Skype Translator and, although we’ve built a solid foundation, we know that in many respects our work has just begun.