

Do you have a
VIP
(Very Inclusive Product)
for Unified Comms?



Introduction

Inclusive practice has become a defining trait in business in recent years. The goal to redefine structures and processes to make them more accessible for marginalized groups has seen new rules and regulations introduced across the globe.

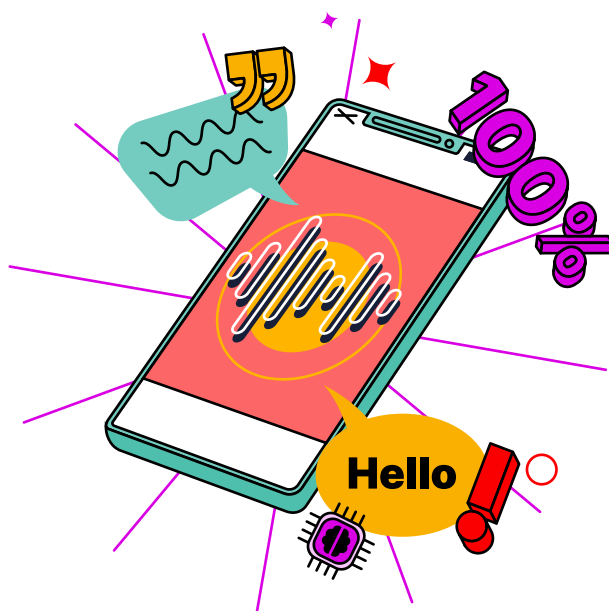
Adhering to accessibility requirements must be front and center with efforts to provide a VIP (very inclusive product) driving your product's development.

For business product leaders, this means ensuring your unified communications offering – whether that's a meeting platform or call center service – can deliver an exceptional user experience, capturing voices no matter how or where they take place. Whether a customer, partner, or internal business conversation is taking

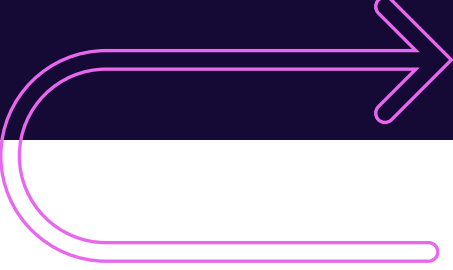
place on the train, in a noisy office, or at home, the ability to accurately transcribe speech in challenging audio environments is paramount for your customers.

But it's not just about where conversations are happening. Your solution must also have the capacity to transcribe all voices – accounting for different languages, dialects, accents, ages, genders, and socio-economic backgrounds – to deliver an inclusive product that stacks up against accessibility regulations, today and tomorrow.

In this Inclusivity Guide, we'll discuss the opportunities for your product's speech-to-text offering – from being part of the roadmap for the APIs you're using to providing customers with the best user experience – with inclusive, accurate, accessible practice at the heart of development.



Changing the Game with **Speechmatics'** Ursa Generation Models



Accuracy is the most important result your speech partner will provide. Your product itself might be ground-breaking, but if the speech tech in your stack can only deliver in certain conditions, for particular voices, then your total addressable market (TAM) is limited from the start, and you will struggle to differentiate against your competitors.

For your speech-to-text offering to be inclusive, it must be accurate across the board. A truly inclusive approach requires more than just the ability to accurately account for variations in accent. Datasets have historically underrepresented those from particular genders, ages, and socio-economic cohorts¹. The result is a general tendency for speech-to-text systems to favor those from specific backgrounds.

It's clear that you can't build inclusive, global products if you don't have a highly accurate Speech API. That's why Speechmatics' Ursa has been built to understand every voice. Which means accurate transcriptions, regardless of the demographic of the speaker, their dialect, accent, or language.

Our latest release, the Ursa generation, is our most inclusive yet. Ursa has a relative accuracy 24% higher than the next leading speech-to-text provider when transcribing English for International speakers (see Figure 1).



24% higher than the next leading speech-to-text provider



Word error rate (WER) is the most common metric for measuring and comparing the accuracy of speech-to-text models. Broadly speaking, it is the number of errors in a transcript divided by the total number of words transcribed. So, the lower the WER the more accurate the model.

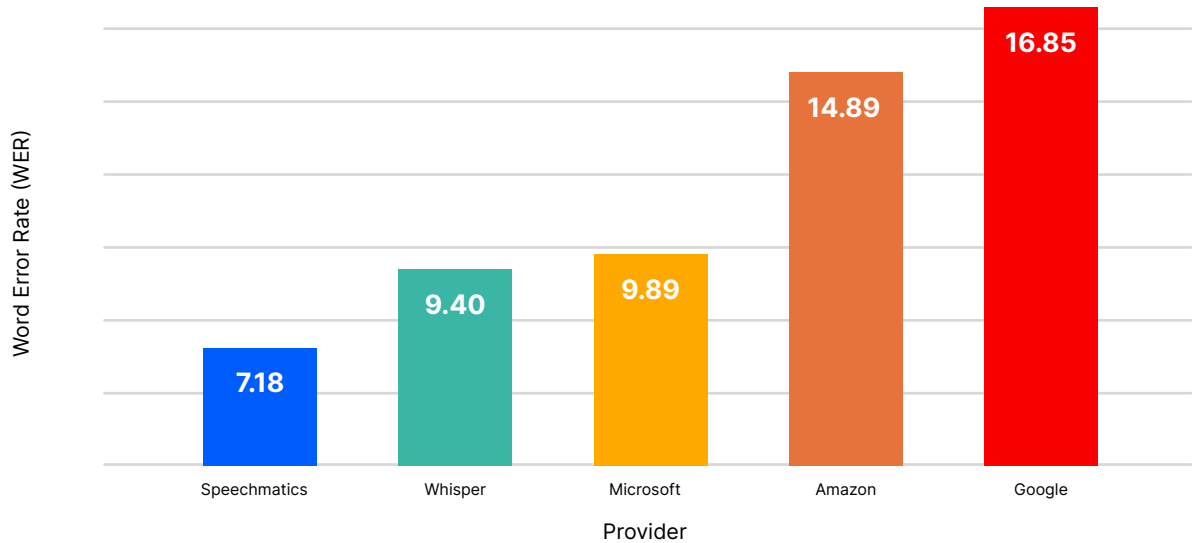


Figure 1. Word error rates (WERs) assessed on 26 hours of speech from speakers across the globe with varied accents, comparing against several speech-to-text vendors, with lower WERs indicating better performance (results based on the evaluation set of Common Voice Dataset²).

We put our Ursa generation models through their paces, evaluating transcription performance against other providers using a combination of Casual Conversation³ and CORAAL⁴ datasets. Whatever the age, gender identity, socio-economic status, and education level of the speakers, Ursa delivered a 10% relative increase in accuracy across all demographics. Giving you confidence that, with Speechmatics, your product can deliver for a wide range of speakers.

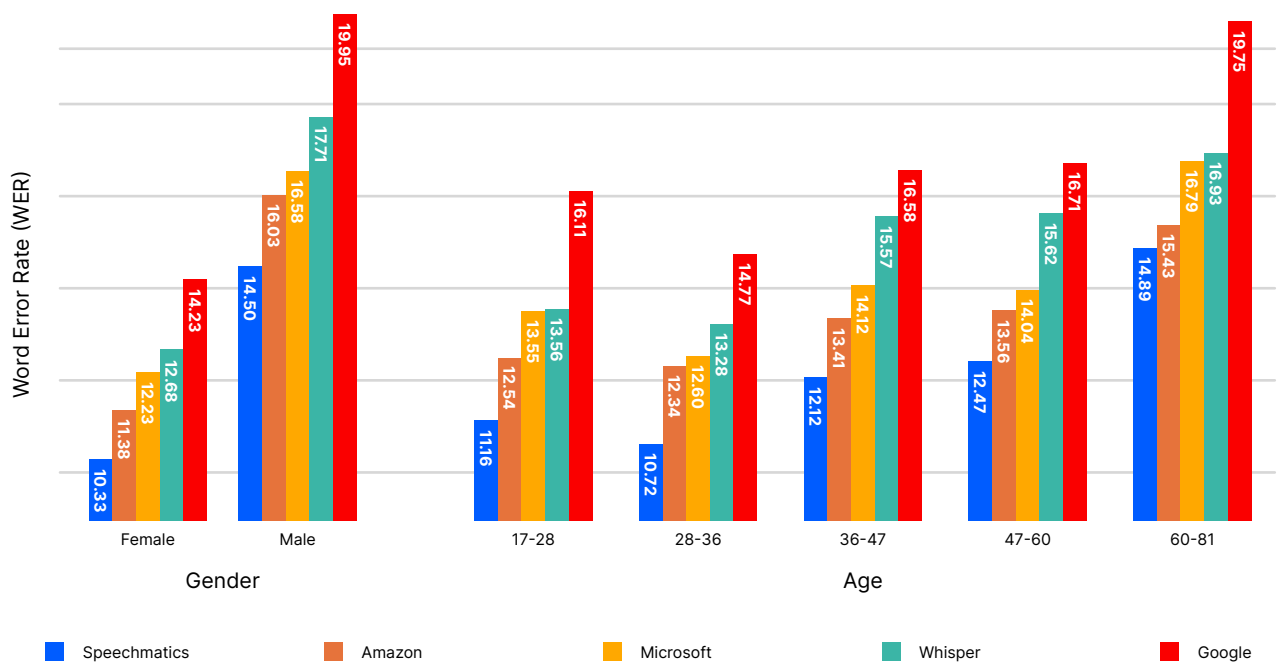


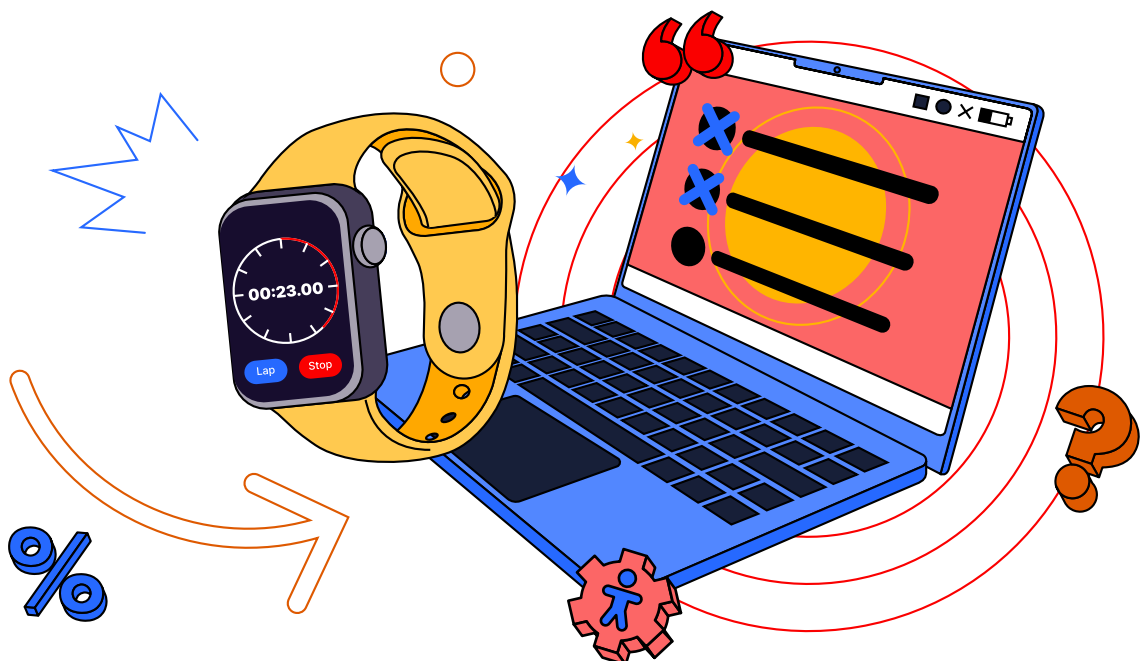
Figure 2. Ursa is consistently the most accurate speech-to-text system across age and gender, with a 30% and 25% relative lead on male and senior voices compared to Google respectively.

We're obsessed with your end user. So, **what are you thinking about?**

Prioritizing user experience is a proven strategy for building successful, enduring products that can deliver a differentiated experience and long-term growth in consumption.

For great speech technology, that means delivering without being noticed. Users shouldn't need to consider background noise or moderate their speech, they should be able to rely on flawless transcriptions regardless of who is speaking, the language they're using, or the environment they're in.

You need to know your speech partner will deliver an exceptional end-user experience, whether the speaker is in a taxi, on a train platform, in a busy office, or working from home. We tested Ursa against some of the most challenging audio conditions using an AVICAR dataset⁵, and our models delivered a 33% relative lead over the closest competitor.



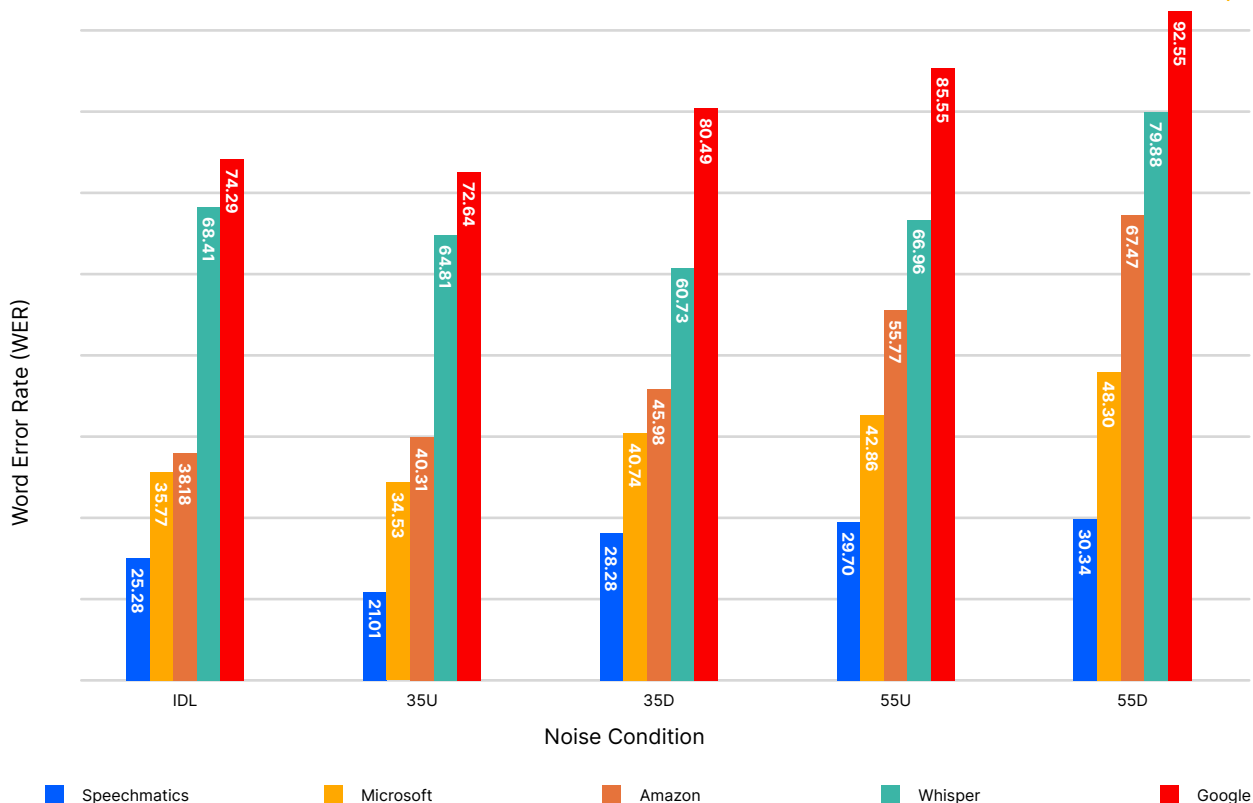


Figure 3. While the word error rate for most providers tends to increase as noise increases, Ursa remains consistently accurate across varying levels of noise with a relative lead spanning between 30% and 40%.

An exceptional user experience requires a comprehensive feature set. Speechmatics' commitment to your end user means our API provides every feature you need to deliver an unparalleled experience.

Transcripts are fully searchable post-call, with speaker and channel diarization capturing exactly what was said by each participant, even when they talk over each other. Our high accuracy also minimizes the need for human transcription or note-taking – an important time saver given the number of meetings now conducted remotely.

Speechmatics' API also covers the spoken languages of over half the world's population, ensuring accessibility for users while maximizing the TAM for your product. And because Speechmatics is committed to excellence, we deliver the same accuracy – whether the speaker is in a busy contact center, on a train, or sitting in a home office – so users can rely on your tech, wherever they are.

Today, your home markets. Tomorrow, the world.



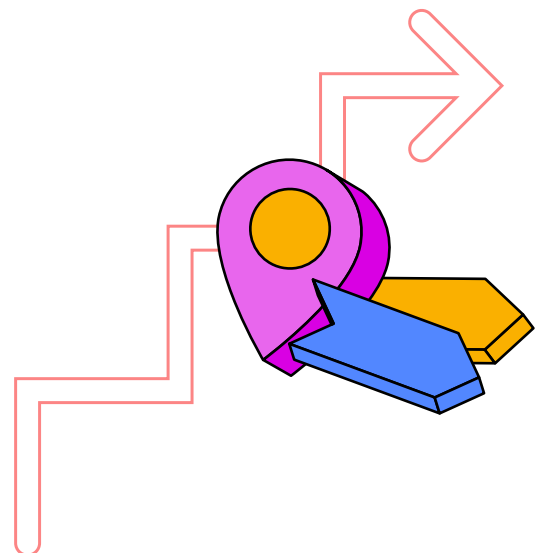
As a product leader, you know the right speech partner is critical to your global expansion. Speechmatics' API covers 48 languages with unrivaled accuracy, so you can be confident that you'll have the language coverage you need, when you need it.



***maintains a
22% relative
lead***

For product teams, Speechmatics gives you the foundation to scale quickly into new markets. Our API provides certainty that your product will be inclusive of not only different languages but also a range of accents and dialects, from Australian English to Brazilian Portuguese.

Our focus on inclusion ensures our partners can deliver across a wide variety of industries and for speakers from all backgrounds. Ursa has been built to understand every voice, and maintains a 22% relative lead over the rest of the market when it comes to often-overlooked dialects, such as African American Vernacular English.



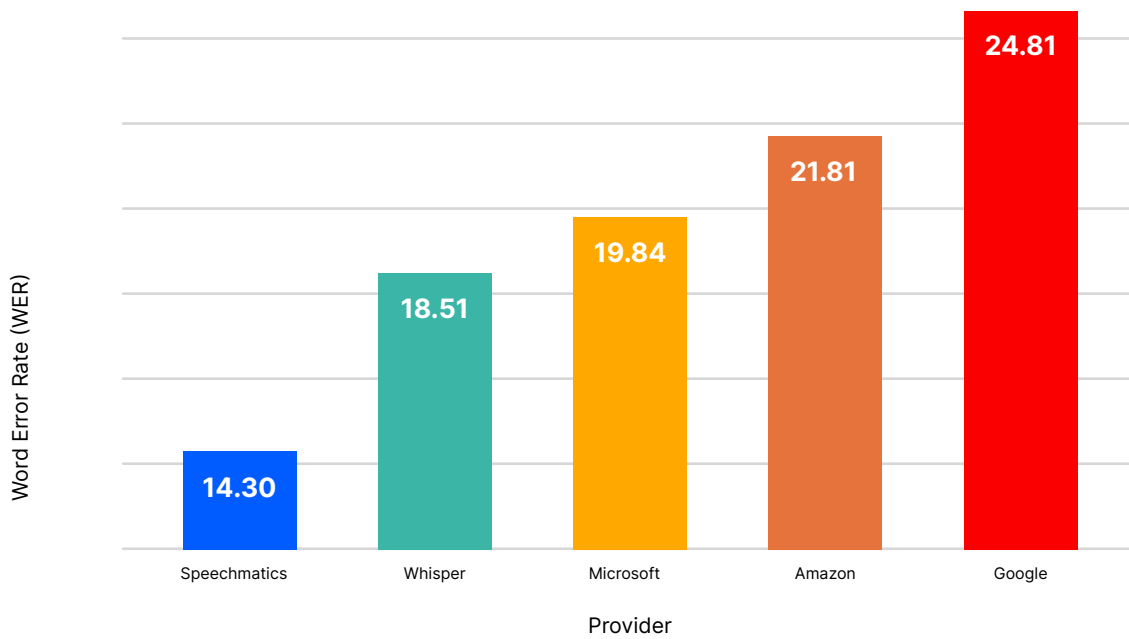
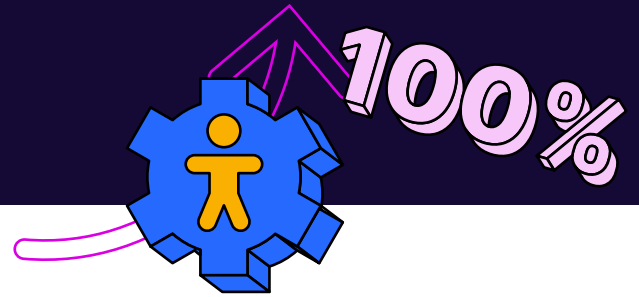


Figure 4. WERs for over 100 hours of African American Vernacular English conversation (results based on CORAAL dataset®).

Our comprehensive feature set and ongoing development means you can continue to innovate and deliver for a wide range of customers, reducing time to market and supporting strategic entry into new regions with flexible deployment options, a choice of transcription modes, custom dictionaries, speaker labels, word timings, confidence scores, and much more.



Our product is built for everyone. **Is yours?**



Today's landscape for Unified Communications products is complex. Uncertainty hides around every corner, from economic turbulence to political instability worldwide, and is influencing the market in unpredictable ways.

As a product leader, you know that ensuring your product stacks up against tomorrow – whatever it might bring – is a vital component for your product's future success.

And the ability to be quick to market means your product needs the right capabilities at the right time. Which also means an assurance that the tools that make up your product are ready and willing to move with you to take advantage of any opportunities and outstrip competitors.

While there's a tendency to opt for big tech because it's well-known, when it comes to speech-to-text, big tech doesn't mean big voice capabilities. In fact, as our results have shown, it typically hampers your ability to deliver a robust, wide-ranging voice-capturing solution as part of your product.

If you want a reliable product that can scale, compete, respond quickly to the needs of inbound customers as well as changing market expectations, and deliver an exceptional experience for end users – whoever they might be – you need Speechmatics.

Here's why:

- When compared with Microsoft, Amazon, and Google, Speechmatics' Ursa delivered a 10% relative increase in accuracy for transcriptions across all demographics. So, no matter the gender identity, socio-economic status, age, or education level of speakers you can be confident your Unified Communications offering can capture all voices in any business environment
- In challenging audio conditions – from noisy call centers to home offices – our Ursa generation models delivered a 33% relative lead over the closest, big tech competitor
- Speechmatics' Ursa maintained a 22% relative lead over the rest of the market when transcribing dialects that are often overlooked. Meaning, you can be certain you're delivering an inclusive speech-to-text offering to market that helps your product adhere to accessibility regulations

Can you say your product
is a VIP (Very Inclusive Product)?

Does your speech-to-text offering
advance your product while
also helping you adhere to the
latest regulations?

Get in touch with us today and
discuss how you can take your
product's speech-to-text offering to
the next, most inclusive level.

Get in touch.

Appendix

1. Liu, Chunxi, et al. "Towards measuring fairness in speech recognition: casual conversations dataset transcriptions." ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.
2. Ardila, Rosana, et al. "Common voice: A massively-multilingual speech corpus.", 2019.
3. Liu et al., "Towards Measuring Fairness in Speech Recognition: Casual Conversations Dataset Transcriptions," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.
4. The Online Resources for African American Language Project.
<http://oraal.uoregon.edu/coraal>.
5. <http://www.isle.illinois.edu/sst/AVICAR/>
6. The Online Resources for African American Language Project.
<http://oraal.uoregon.edu/coraal>

