

The Speechmatics Approach to Global Spanish.

Accent-Independent
Speech Recognition.

October 2020

WHITEPAPER



Diversity of the Spanish Language

With approximately 500 million speakers globally, Spanish is the second most natively spoken language in the world, and fourth most spoken language overall.

Figure 1 shows the regions and number of individuals that speak Spanish worldwide. Because of its global appeal and diversity, Spanish poses a significant challenge when it comes to providing consistent and accurate speech-to-text.



Figure 1: Spanish Speakers Worldwide

	Spanish Dialect	Description	Approximate Speakers Worldwide	% of Worldwide Speakers
The Americas	Latin American	This is the dialect of urban mainland Mexico, Colombia, Peru, Bolivia, and the majority of Central and South American countries.	321,890,586	64.7%
	Rioplatense	This dialect is spoken in the River Basin region between Argentina and Uruguay, as well as in both countries.	49,728,234	10.0%
	Caribbean	This dialect is spoken in Cuba, Puerto Rico, the Dominican Republic, and along the East coast of Mexico and Central America.	25,017,889	5.0%
	United States	Mostly Latin American from Mexico and other parts of Central America.	50,000,000	10.0%
Peninsular Spanish (Spain) Spanish from Spain (aka Castilian Spanish, Iberian Spanish, Peninsular Spanish, or European Spanish)	Castilian	This term applies to the official Spanish language, spoken in Northern and Central Spain.	47,737,941	9.6%
	Andalusian	This dialect, spoken in Southern Spain, is the second-most popular in the country after Castilian.		
	Murcian	This dialect is spoken in the Autonomous Region of the Community of Murcia in the South East of Spain.		
Canary Islands	Canarian	The dialect of the Spanish Canary Islands closely resembles the Caribbean Spanish dialect.	1,600,000	0.3%
Gibraltar	Llanito	Llanito is a combination of Andalusian Spanish and British English.	1,722,254	0.31%
Total			497,696,904	100%

Source 1, Source 2

Approximately 90% of Spanish spoken in the world is in the United States, Mexico, Central and South America, with the remaining 10% in Spain. According to [CMS.gov](https://www.cms.gov), Spanish was the top non-English language in all but four states in the United States.

The Beginning of the Global Approach

Historically, to get the most accurate results from speech recognition technology, providers were required to build accent-specific language packs such as Mexican-Spanish. Confronted with different accents, dialects and other regional variations in speech, this approach has become hard to scale globally as the world becomes more connected and diverse.

Speechmatics changed its approach to building and adapting languages in 2018, with the introduction of [Global English](#). As the first and only company to do away with accent-specific language packs, Speechmatics' global approach supports many accents and dialects in a single language pack.

This approach is optimized for real-world usability. The introduction of [Global English](#) optimizes deployment and provides operational and management simplicity for users while outperforming competitor offerings which are designed for specific accents and dialects.

[ORIGINAL LINK BROKEN](#)

Global Languages Redefined

The diversity of Spanish accents and dialects makes supporting it as a language for use in speech-to-text transcription a significant challenge. To get accurate transcripts, providers have created multiple Spanish language packs specializing in a specific region or speaker profile. While this offers a solution to the challenge presented by multiple accents and dialects, it is not scalable nor the best solution for real-world use cases and applications.

With such a variety of accents and dialects within Spanish, it is not enough for companies to select a specialized language pack. In the real world, audio files often include more than one speaker from multiple regions, all with different accents, dialects and idiosyncrasies. Deploying accent-specific language packs requires organizations to make a best guess as to the best language pack to use for each audio file. It also requires them to host and store multiple language packs for one language, adding to operational complexities and costs for what should be an efficient, automated process and workflow.

[ORIGINAL LINK BROKEN](#)

Spanish in Contact Centers

In a [2019 report](#) conducted by Speechmatics, 82% of contact center companies expressed languages to be important with Spanish and English being the most important language for their contact center.

As brands look to grow their reach, they also have to meet customer expectations to optimize their experience, drive loyalty and reduce churn. This means delivering localized and personalized services to those customers. The ability to use any-context speech recognition technology to transcribe Spanish accurately enables contact centers to use voice data to improve customer experiences and empower agents.

A 2016 [survey by ICMI](#), discovered that 57% of customers expect the service from their contact center to be in their native language – as opposed to the primary language of the contact center.

Global Spanish Competitor Comparison

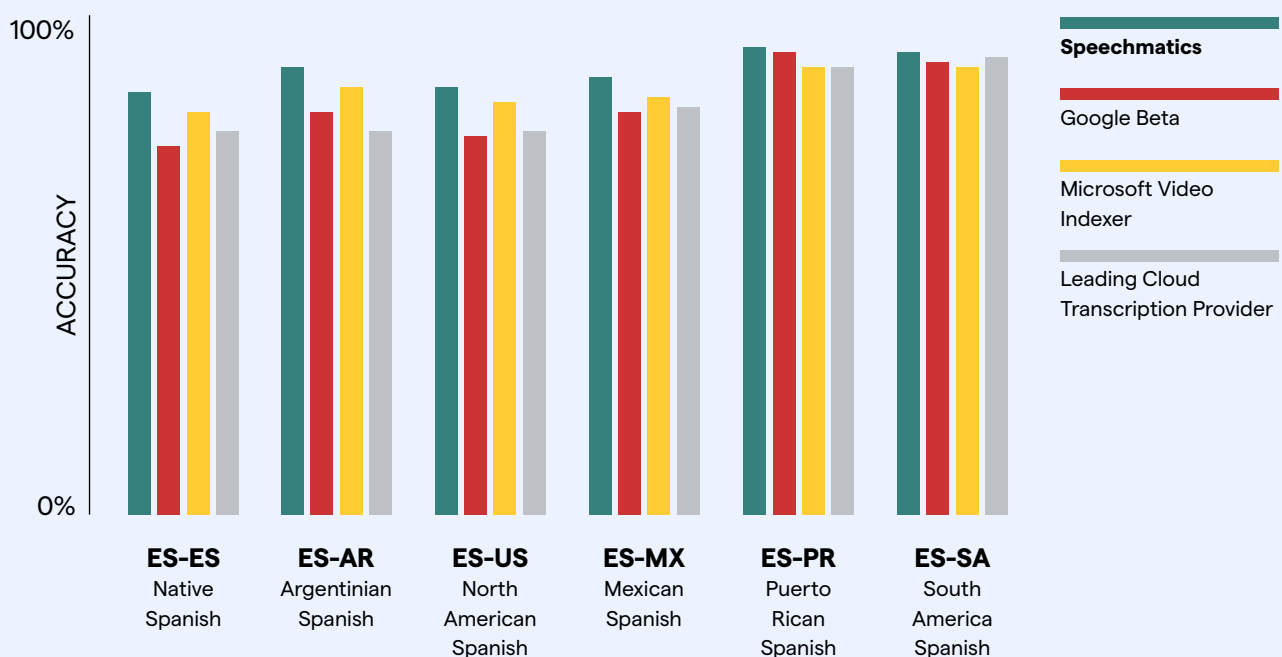
Proving the Theory

To test our global approach, we created a Global Spanish language pack leveraging the knowledge gathered from [Global English](#), utilizing the latest advancements in machine learning and applied proprietary language training techniques. We tested Global Spanish against other ASR providers, figure 2 shows the results.

Figure 2: One Model to Rule Them All

Spanish - Regional Accuracy Testing

Figures taken from August 2020. Test sets comprised of almost 8.5 hours of diverse audio and transcribed text covering multiple use cases. Accented test files included variations in gender, age, region and ethnicity of speakers.



The testing in figure 2 was done using Speechmatics' own audio files. This provides representative data on how Speechmatics' any-context speech recognition engine compares to other ASR providers. Accuracy percentage is calculated by $100 - \text{word error rate (WER)}$. No consideration is made to the accuracy of punctuation or any other factor that might contribute to whether one transcription provider is more accurate or better than another.

The values in figure 2 are representative of the data that Speechmatics used in testing only. The Speechmatics test sets comprise of pre-recorded, real-world audio data from audio files submitted by our partners and customers with their permission given to be used for this testing.

The Results

Our testing found that for each variation of Spanish, Speechmatics' Global Spanish language pack was the best option in every instance.

[ORIGINAL LINK BROKEN](#)

Real-World Value

Ease of Use

This single, multi-use solution means users do not need to identify which Spanish variant is being spoken. When audio files feature multiple speakers with different accents, or where speaker accents are not known in advance, Global Spanish provides reliable results over a broader range of speakers.

Fewer Models to Maintain and Update

By focusing resources on maintaining and updating fewer language packs, Speechmatics can increase quality, improve accuracy and ensure reliability of fewer packs for our customers and partners.



Innovating to Deliver Better Performance

Improved Algorithms

Speech recognition has advanced hugely in recent years, giving step change improvements in a field used to marginal gains. In particular, modern neural network architectures are capable of generalizing across variations in speech. Deep neural networks feature multiple layers between input and output. This effectively gives us the performance of a variety of specialized models, all in one comprehensive language pack.

Greater Computing Power

Single modern servers are more powerful than old, room-filling supercomputers. This astonishing rise in compute power, coupled with repurposing of GPUs gives masses of computing power. The advancements in compute power allows Speechmatics to train models based on more data, capable of supporting more variations in a single language pack.

Data Availability

By investing more time into gathering data from a wide range of sources, we have created a huge and diverse training corpus. This allows us to train models with a much wider range of applications than ever before.

Speechmatics is already delivering leading levels of accuracy through traditional approaches to speech recognition. However, we are also investing in research and development to find new ways of solving problems to help our customers and partners innovate with voice. These approaches will deliver even better levels of accuracy across more languages while making it easier to own and operate the Speechmatics solution.

Future-Proofing

Speechmatics frequently tests and benchmarks against other providers to ensure we're delivering the best speech recognition in the market. By defying the industry convention and producing accent-independent language packs, Speechmatics is on a journey to make all languages truly global.

By following our [success with Global English](#), Global Spanish is fast, accurate, reliable and now more flexible, convenient and inclusive. Global Spanish offers users speech recognition for the future.



