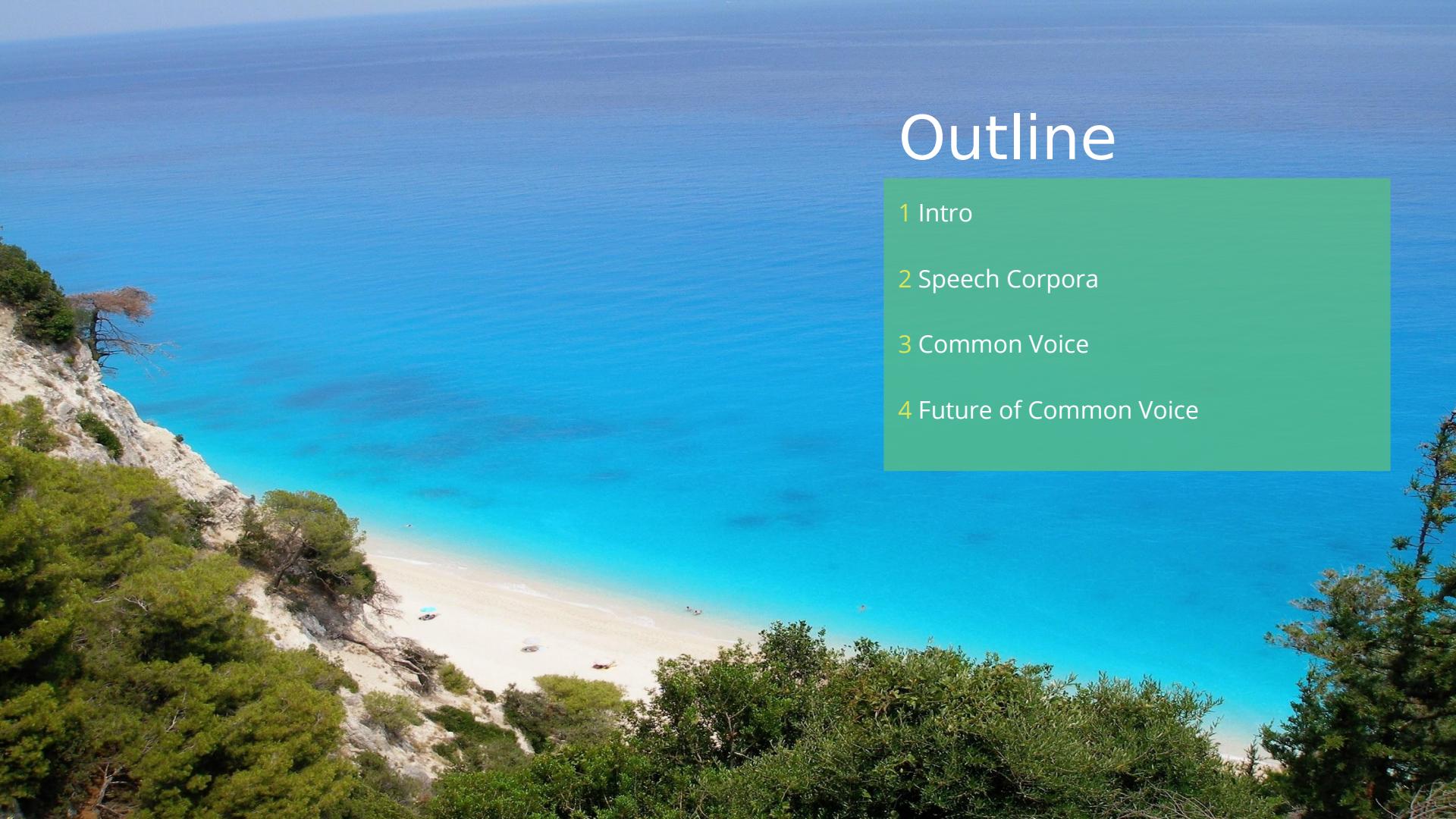


Mozilla's Common Voice



The background of the slide features a wide-angle photograph of a beautiful coastal scene. On the left, a light-colored cliff face with sparse greenery descends towards a white sandy beach. The water is a vibrant turquoise color, transitioning to a darker blue further out. The sky above the water is clear and bright.

Outline

1 Intro

2 Speech Corpora

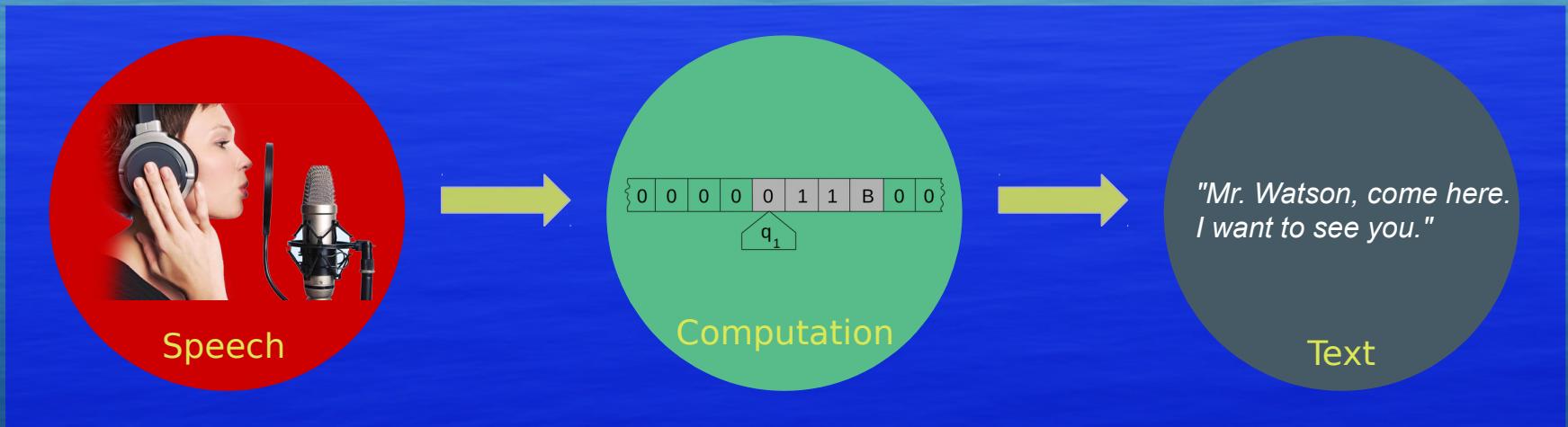
3 Common Voice

4 Future of Common Voice

1 Intro



Intro: Speech-to-Text



Intro: Speech Corpora



Speech

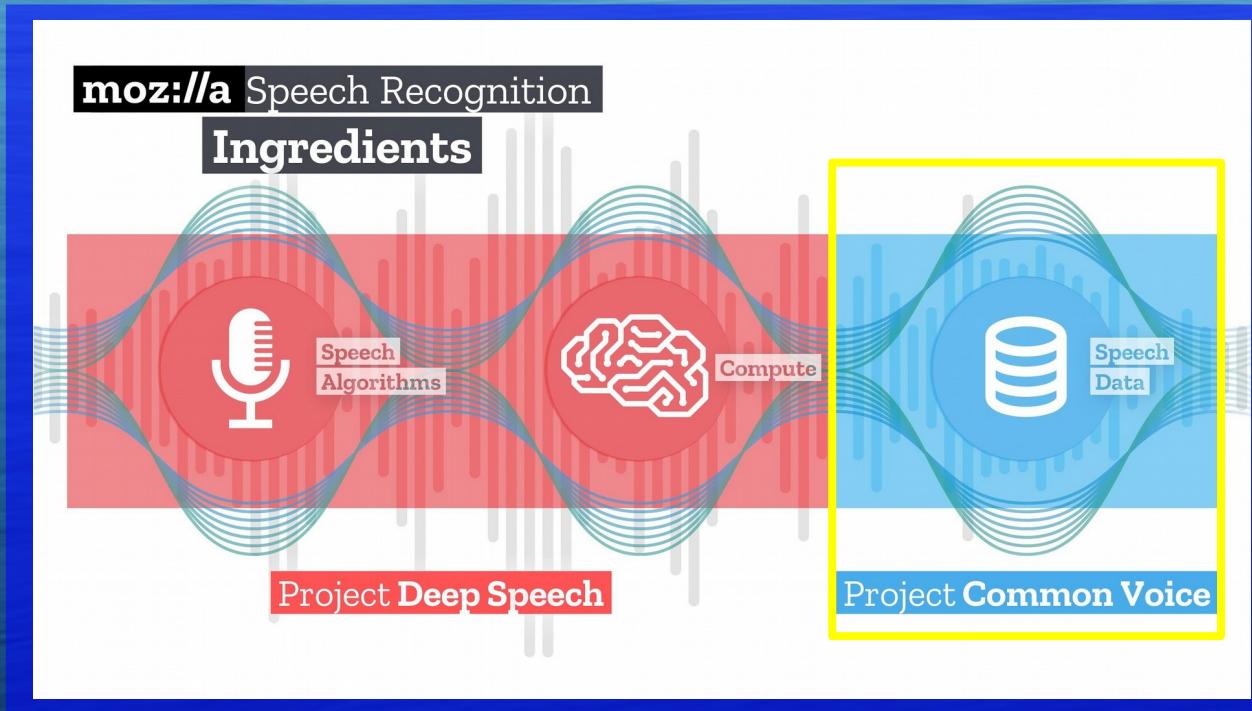


Speech



Speech

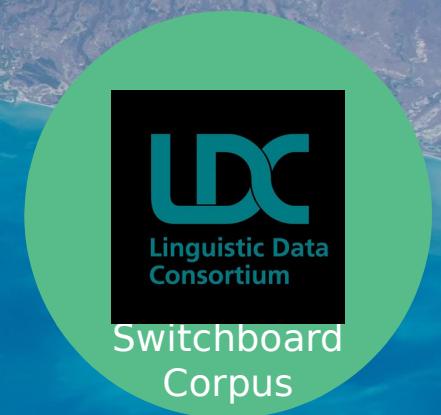
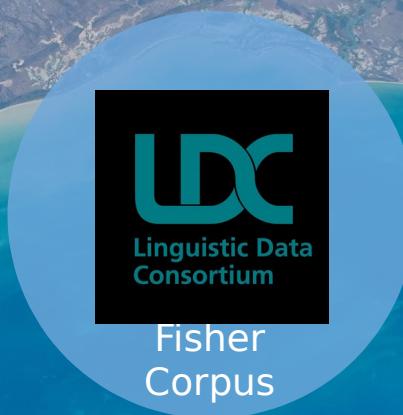
Intro: Common Voice

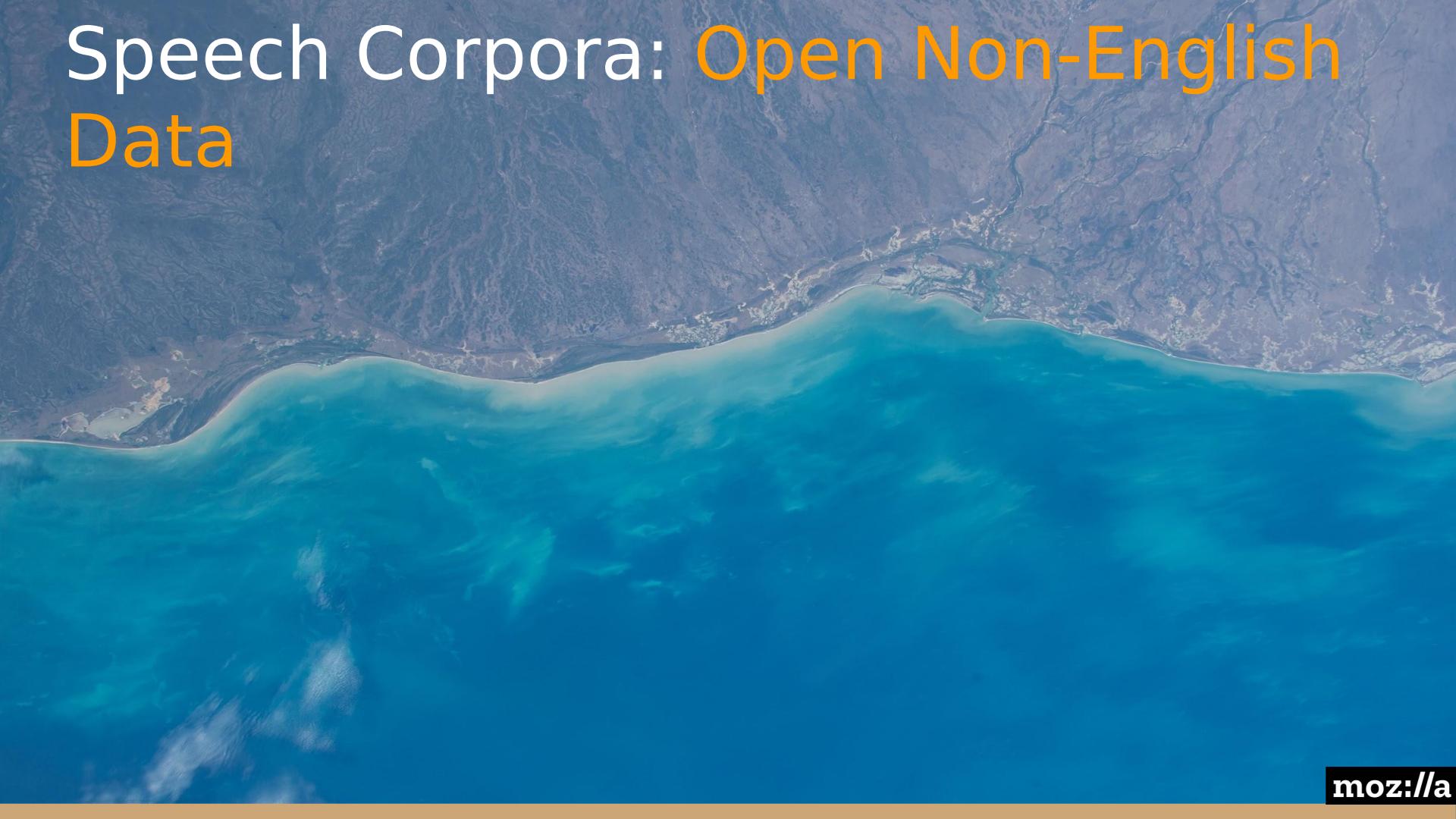


2 Speech Corpora



Speech Corpora: English Data



The background image is a high-angle aerial photograph of a coastal region. The left side shows a dark, hilly or mountainous area with a network of winding roads and small settlements. A sharp, light-colored coastal strip runs along the middle, separating the land from a vast expanse of bright turquoise blue water. The water has some darker, wavy patterns where it meets the shore.

Speech Corpora: Open Non-English Data

Speech Corpora: Open Non-English Data



3

Common Voice



The background image shows a vast desert landscape with a range of mountains in the distance. In the foreground, there are large, light-colored sand dunes with dark, wavy patterns. The sky is clear and blue.

3.1 Frontend

Frontend: Homepage

The screenshot shows the Mozilla Common Voice homepage. At the top, there's a navigation bar with links for 'CONTRIBUTE', 'DATASETS', 'LANGUAGES', and user account options ('Log In / Sign Up' and language selection). Below the navigation, there are two main sections: 'Speak' on the left and 'Listen' on the right. Each section features a large circular icon with a microphone symbol (red for Speak, green for Listen) and a descriptive subtitle. Below these are two large, semi-transparent audio waveform graphics in red/pink and blue/cyan respectively. At the bottom, there are two paragraphs of explanatory text and a 'READ MORE' button.

Common Voice
moz://a

CONTRIBUTE DATASETS LANGUAGES

0 0 Log In / Sign Up EN

Speak Donate your voice

Listen Help us validate voices

Common Voice is Mozilla's initiative to help teach machines how real people speak.

Voice is natural, voice is human. That's why we're excited about creating usable voice technology for our machines. But to create voice systems, developers need an extremely large amount of voice data.

Most of the data used by large companies isn't available to the majority of people. We think that stifles innovation. So we've launched Common Voice, a project to help make voice recognition open and accessible to everyone.

[READ MORE](#)

Frontend: Speak

The screenshot shows a user interface for a speech recognition or recording application. At the top, there's a navigation bar with a back arrow, the word "Speak" underlined in blue, and "Listen". To the right, it says "1/5 Clips". Below the navigation is a large text input field containing the sentence "I wound up the coil of wires.". Above this text field, instructions say "Click then read the sentence aloud". A large red microphone button is centered below the text field. At the bottom, there are several buttons: "Shortcuts" (with a keyboard icon), "Unable to speak right now?", "Skip >>", and "SUBMIT".

1/5 Clips

START RECORDING 1

2

3

4

5

I wound up the coil of wires.

Click then read the sentence aloud

Speak Listen

Shortcuts

Unable to speak right now?

Skip >>

SUBMIT

Frontend: Listen

The screenshot shows a user interface for a transcription or listening exercise. At the top, there are two tabs: "Speak" and "Listen", with "Listen" being underlined, indicating it is the active mode. In the top right corner, it says "1/5 Clips". Below the tabs, a button with a speaker icon has the number "1" next to it, with other numbers 2, 3, 4, and 5 listed vertically to its right. A central text area contains the sentence: "So, he asked if the old man's blessing was still with him." Above this text, a small instruction reads: "Click ▶ did they accurately speak the sentence?" Below the text area is a large circular button containing a green play/pause icon. To the left of this button is a thumbs-up icon with the word "YES" and to the right is a thumbs-down icon with the word "NO". At the bottom left is a "Shortcuts" button, and at the bottom right is a "Skip >>" button.

1/5 Clips

1

2

3

4

5

Click ▶ did they accurately speak the sentence?

So, he asked if the old man's blessing was still with him.

YES

NO

Shortcuts

Skip >>

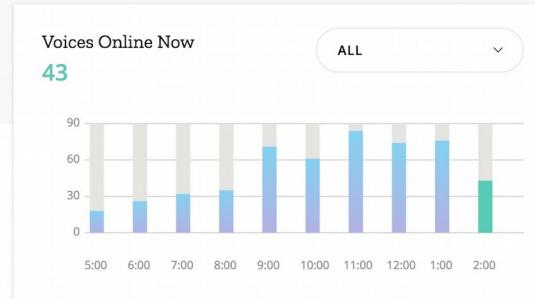
Frontend: General Statistics

Common Voice is Mozilla's initiative to help teach machines how real people speak.

Voice is natural, voice is human. That's why we're excited about creating usable voice technology for our machines. But to create voice systems, developers need an extremely large amount of voice data.

Most of the data used by large companies isn't available to the majority of people. We think that stifles innovation. So we've launched Common Voice, a project to help make voice recognition open and accessible to everyone.

[READ MORE](#)



Frontend: Datasets

The screenshot shows the homepage of the Common Voice website. At the top, there's a navigation bar with links for 'CONTRIBUTE', 'DATASETS' (which is highlighted in red), and 'LANGUAGES'. On the right side of the header are icons for user profile ('KDAVIS'), language ('EN'), and a dropdown menu. Below the header, a large button with a download icon and the text 'Download Common Voice Data' is centered. The main content area features a paragraph about the project's goal: 'We are building an open and publicly available dataset of voices that everyone can use to train speech-enabled applications.' Below this, another paragraph explains the project's mission: 'We believe that large and publicly available voice datasets foster innovation and healthy commercial competition in machine-learning based speech technology. This is a global effort and we invite everyone to participate. Our aim is to help speech technology be more inclusive, reflecting the diversity of voices from around the world.' A link 'Get Started with Speech Recognition' is provided. Further down, there's a section titled 'Other voice datasets...' with cards for 'LibriSpeech', 'TED-LIUM Corpus', 'VoxForge', and 'Tatoeba'. Each card contains a brief description, the dataset name in bold, and a 'Download Data' button.

Common Voice
moz://a

CONTRIBUTE DATASETS LANGUAGES

0 35 KDAVIS EN

Download Common Voice Data

We are building an open and publicly available dataset of voices that everyone can use to train speech-enabled applications.

We believe that large and publicly available voice datasets foster innovation and healthy commercial competition in machine-learning based speech technology. This is a global effort and we invite everyone to participate. Our aim is to help speech technology be more inclusive, reflecting the diversity of voices from around the world.

Get Started with Speech Recognition

Other voice datasets...

LibriSpeech
LibriSpeech is a corpus of approximately 1000 hours of 16Khz read English speech derived from read audiobooks from the LibriVox project.
License: CC-BY-4.0

[Go to LibriSpeech](#)

TED-LIUM Corpus
The TED-LIUM corpus was made from audio talks and their transcriptions available on the TED website.
License: CC-BY-NC-ND 3.0

[Download Data](#)

VoxForge

Tatoeba

Frontend: Profile Sign-Up

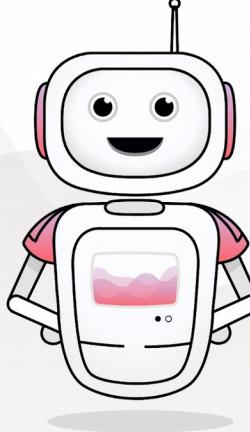
Help us build a high quality, publicly open dataset

Having a profile is not required to contribute though it is helpful

[Sign up for an account](#)

BENEFITS **WHAT'S PUBLIC?**

1. Make your submitted data as rich as possible by providing some anonymous demographic data. We de-identify all demographic data before making it public.
2. Profile information improves the audio data used in training speech recognition accuracy.
3. Keep track of your progress and metrics across multiple languages.
4. See how your progress compares to other



Optional Demographic Data

Age: Select

Sex: Select

Frontend: Dashboard

The screenshot displays the Mozilla Common Voice dashboard with the following sections:

- Stats:** Shows progress towards recording 1,200 clips (0/5) and validating 2,400 clips (35/50).
 - Clips You've Recorded:** 0 / 5. Progress bar: 0% (red).
 - Clips You've Validated:** 35 / 50. Progress bar: 70% (green).
- Contribution Activity:** A chart showing recorded clips over time (6:00 to 3:00). The chart shows a constant rate of approximately 1 clip per hour.
- Top Contributors:** A table ranking contributors by recorded and validated clips.

Rank	User	Recorded Clips	Validated Clips	Completion %
01	Rafik	15600	15498	99.36 %
02	Leparleur	5300	5192	97.96 %
03	Patrick Lisiecki	5243	5016	95.67 %
04	ubahnverleih	4742	4450	93.84 %

Frontend: Profile: Demographics

The screenshot shows the 'Profile' section of the Common Voice website. At the top, there's a navigation bar with links for 'CONTRIBUTE', 'DATASETS', and 'LANGUAGES'. On the right side of the header, there are icons for a microphone (0 recordings), a play button (35 datasets), a user icon ('KDAVIS'), and language selection ('EN').

The main content area is titled 'Profile' and contains several sections:

- Why does this matter?**: A note explaining that providing personal information helps improve speech recognition accuracy.
- User Name**: kDavis
- Leaderboard Visibility**: Hidden
- Age**: 40 - 49
- Sex**: Male
- Native Language**: English
- Accent**: United States English
- Additional Language**: Deutsch
- Accent**: Amerikanisches Deutsch

At the bottom, there's a 'Save' button and a 'Add Language' button with a plus sign. The footer features icons for a globe, a cat, a speech bubble, and a share symbol.

Frontend: Profile: Goals

The screenshot shows the 'Goals' section of the Common Voice profile. On the left, a sidebar lists 'Profile', 'Avatar', 'Goals' (which is selected), 'Settings', 'Delete Profile', and 'Download My Data'. The main area displays three sections: 'Streaks', 'Speak', and 'Listen'. Each section has five cards representing different goal levels.

Section	1	3	5	10	15
Day	1 Day	3 Days	5 Days	10 Days	15 Days

Section	5	20	50	100	150
Recordings	5 Recordings	20 Recordings	50 Recordings	100 Recordings	150 Recordings

Section	5	20	50	100	150
Validations	5 Validations Jun 29, 2018	20 Validations Jun 29, 2018	50 Validations	100 Validations	150 Validations

Frontend: Profile: Delete Profile

The screenshot shows the 'Profile' section of the Common Voice website. At the top, there's a navigation bar with links for 'CONTRIBUTE', 'DATASETS', and 'LANGUAGES'. On the right side of the header, there are icons for a microphone (0 recordings), a play button (35 datasets), and a user account named 'KDAVIS'. Below the header, a sidebar on the left lists several options: 'Profile', 'Avatar', 'Goals', 'Settings', 'Delete Profile' (which is highlighted with a blue background), and 'Download My Data'. The main content area contains a question: 'Would you like to request your voice recordings be deleted too, or do you prefer to keep them in the Common Voice dataset?*'. Below this question are two buttons: 'Remove' (gray) and 'Keep' (green). To the right of the question, a note states: '* Indicates required field'. Further down, a blue button labeled 'Delete Profile' is visible. At the very bottom of the page, there are five small icons: a globe, a cat, a speech bubble, a gear, and a right-pointing arrow.

Common Voice
moz://a

CONTRIBUTE DATASETS LANGUAGES

0 35 KDAVIS EN

Profile Avatar Goals Settings Delete Profile Download My Data

Would you like to request your voice recordings be deleted too, or do you prefer to keep them in the Common Voice dataset?*

* Indicates required field

Remove Keep

Your anonymous voice recordings will remain in the Common Voice dataset. Once you delete your profile you will no longer be able to submit a request to remove your recordings from the dataset.

Delete Profile

Frontend: Profile: Download My Data

The screenshot shows the Common Voice profile page. At the top, there's a navigation bar with links for 'CONTRIBUTE', 'DATASETS', and 'LANGUAGES'. On the right side, there are icons for a microphone (0 recordings), a play button (35 datasets), a user icon (KDAVIS), and language selection (EN). A sidebar on the left contains links for 'Profile', 'Avatar', 'Goals', 'Settings', 'Delete Profile', and 'Download My Data'. The main content area asks if the user wants to request their voice recordings be deleted from the dataset. A Firefox file dialog is overlaid, showing the file 'profile.txt' has been chosen to open. The dialog also asks what Firefox should do with the file, with 'Save File' selected. To the right of the dialog, a note states: 'Your anonymous voice recordings will remain in the Common Voice dataset. Once you delete your profile you will no longer be able to submit a request to remove your recordings from the dataset.'

Common Voice
moz://a

CONTRIBUTE DATASETS LANGUAGES

0 35 KDAVIS EN

Profile Avatar Goals Settings Delete Profile Download My Data

Would you like to request your voice recordings be deleted too, or do you prefer to keep them in the Common Voice dataset?*

* Indicates required field

Opening profile.txt

You have chosen to open:
profile.txt
which is: Plain Text Document (154 bytes)
from: data:

What should Firefox do with this file?
 Open withTextEdit (default)
 Save File
 Do this automatically for files like this from now on.

Cancel OK

Frontend: Languages...

Common Voice
moz://a

CONTRIBUTE DATASETS LANGUAGES

0 0 Log In / Sign Up EN ▾



Don't see your language on Common Voice yet?
Request a Language

Launched

Search

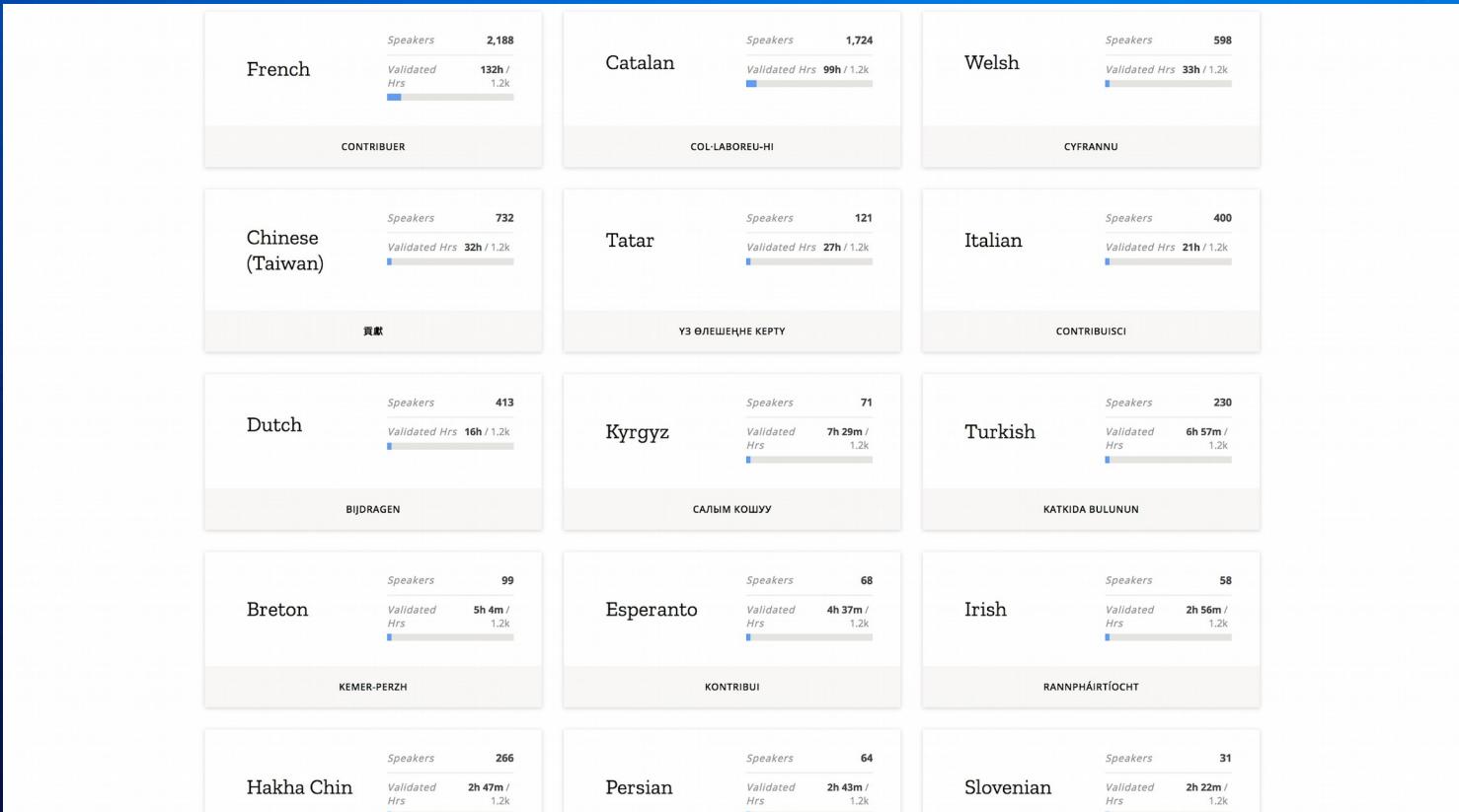
For these launched languages the website has been successfully localized, and has enough sentences collected, to allow for ongoing *Speak* and *Listen* contribution.

Language	Speakers	Validated Hrs
English	35,305	651h / 1.2k
German	3,695	235h / 1.2k
Kabyle	451	162h / 1.2k
French	2,188	132h / 1.2k
Catalan	1,724	99h / 1.2k
Welsh	598	33h / 1.2k

CONTRIBUTE MITMACHEN TTEKKI

CONTRIBUER COL-LABOREU-HI CYFRANNU

Frontend: Languages...



Frontend: Languages...

Hakha Chin

Speakers 266

Validated Hrs 2h 47m / 1.2k

AW TUNNAK

Persian

Speakers 64

Validated Hrs 2h 43m / 1.2k

مشترک

Slovenian

Speakers 31

Validated Hrs 2h 22m / 1.2k

PRISPEVAJ

Estonian

Speakers 54

Validated Hrs 2h 9m / 1.2k

LÖÖ KAASA

Chuvash

Speakers 36

Validated Hrs 1h 16m / 1.2k

XUTSHÄH

Basque

Speakers 22

Validated Hrs 28m / 1.2k

LAGUNDU

See Less

In Progress

In progress languages are currently being built for contribution by our communities; their progress reflects where they are across the website localization and sentence collection phases.

Chinese (China)

Localized 100%

Sentences 2883 / 5000

参与进来

Chinese (Hong Kong)

Localized 73%

Sentences 636 / 5000

加入我們

Spanish

Localized 100%

Sentences 110 / 5000

PARTICIPA

The background image shows a high-angle aerial view of a desert landscape. The terrain is characterized by a series of parallel, reddish-brown ridges and deep, dark blue-grey valleys. The ridges have a distinctively layered or stratified appearance, suggesting geological processes like sedimentary rock formation. The lighting is dramatic, with the ridges catching bright sunlight and the valleys remaining in deep shadow, emphasizing the three-dimensional nature of the terrain.

3.2 Localization

Localization: Pontoon

Teams Projects Contributors Machinery

Sign in

Common Voice

PRIORITY ★★★★☆ DEADLINE — REPOSITORY github.com/mozilla/voice-web RESOURCES Staging site · Production site CONTACT PERSON Peiying Mo [:CocoMo]

66 %

	TRANSLATED	26,880	ALL STRINGS	40,680
FUZZY	0			
WARNINGS	0			
ERRORS	0			
MISSING	13,800		UNREVIEWED	1,915

TEAMS 90 CONTRIBUTORS INFO 1

Filter teams

LANGUAGE ▾	LOCALE	SPEAKERS	LATEST ACTIVITY	PROGRESS	💡
Abkhaz	ab	200,000	5 hours ago	<div style="width: 47%;"></div>	47%
Acehnese	ace	3,648,134	6 days, 3 hours ago	<div style="width: 28%;"></div>	28%
Adyghe	ady	600,000	2 weeks, 2 days ago	<div style="width: 0%;"></div>	0%
Afrikaans	af	8,642,000	2 months, 1 week ago	<div style="width: 0%;"></div>	0%
Albanian	sq	6,715,000	2 days, 20 hours ago	<div style="width: 97%;"></div>	97%
Amharic	am	37,572,000	—	<div style="width: 0%;"></div>	0%
Arabic	ar	304,854,000	3 days, 6 hours ago	<div style="width: 93%;"></div>	93%
Aragonese	an	54,000	—	<div style="width: 0%;"></div>	0%
Assamese	as	16,272,000	—	<div style="width: 0%;"></div>	0%
Asturian	ast	625,000	1 month, 3 weeks ago	<div style="width: 75%;"></div>	75%

Localization: Pontoon: Resources

Teams Projects Contributors Machinery

Sign in

Common Voice French

PRIORITY ★★★★☆ DEADLINE — REPOSITORY github.com/mozilla/voice-web RESOURCES Staging site · Production site CONTACT PERSON Peiyi Mo [:CocoMo]

96 %

	TRANSLATED	437	ALL STRINGS	452
FUZZY	0			
WARNINGS	0			
ERRORS	0			
MISSING	15		UNREVIEWED	2

RESOURCES 2 CONTRIBUTORS PROJECT INFO 1 TEAM INFO 1

Filter resources

RESOURCE ▾	LATEST ACTIVITY	PROGRESS
cross-locale.ftl	8 months ago	<div style="width: 100%;">100%</div>
messages.ftl	22 hours ago	<div style="width: 96%;">96%</div>

The background image shows a high-angle aerial view of a mountainous landscape. The terrain is characterized by deep, narrow valleys and rugged ridges. The colors are predominantly reddish-brown, suggesting a desert or volcanic environment. Sunlight creates sharp shadows in the valleys, emphasizing the three-dimensional nature of the terrain.

3.3 Sentence Collection

Sentence Collection: Homepage

Common Voice
moz://a

Home
How-to
Add
Review
Login

Welcome to the Common Voice Sentence Collector

This is a website where we collect and review sentences for [Common Voice](#).

The Common Voice Sentence Collector has collected 192957 sentences in 20 languages!

[Discourse](#)
[Report Bugs \(GitHub\)](#)
[Privacy](#)
[Terms](#)
[Cookies](#)

Sentence Collection: How To

Common Voice
moz://a

Home

How-to

Add

Review

Login

How to

Login

You can login by entering any available username and choosing a password, that will automatically create an account for you. Once logged in you can select your languages from the profile section.

Add new sentences

- All sentences you submit must be under [Public Domain \(CC-0\) license](#).
- Numbers. There should be no digits in the source text because they can cause problems when read aloud. The way a number is read depends on context and might introduce confusion in the dataset. For example, the number "2409" could be accurately read as both "twenty-four zero nine" and "two thousand four hundred nine".
- Abbreviations and Acronyms. Abbreviations and acronyms like "USA" or "ICE" should be avoided in the source text because they may be read in a way that does not coincide with their spelling. Additionally, there may be multiple accurate readings for a single abbreviation. For example, the acronym "ICE" could be pronounced "I-C-E" or as a single word.
- Punctuation. Special symbols and punctuation should only be included when absolutely necessary. For example, an apostrophe is included in English words like "don't" and "we're" and should be included in the source text, but it's unlikely you'll ever need a special symbol like "@" or "#."
- Foreign letters. Letters must be valid in the language being spoken. For example, "x" is a letter in the Russian alphabet but is never used in English and so should never appear in any English source text.
- Length. Sentences must be 14 words or less.

We prefer natural/conversational sentences. While phonetic diversity and different words in sentences is important, we are trying to make recording sentences as much fun as possible. Therefore it would be great if you could try to keep your sentences as natural/engaging as possible.

Review Sentences

Make sure the sentence meets the following criteria:

1. The sentence must be spelled correctly.
2. The sentence must be grammatically correct.
3. The sentence must be speakable.
4. If the sentence meets the criteria, click the "yes" button on the right.
5. If the sentence does not meet the above criteria, click the "no" button on the right. If you are unsure about the sentence, you may also skip it and move on to the next one.

Sentence Collection: Login

Common Voice
moz://a

Home
How-to
Add
Review
Login

Log in

Choose your username and password to sign up. If you already have a username, please sign in with the existing password.

username (alphanumeric, no email)

password

Submit

Sentence Collection: Add

Common Voice
moz://a

Home
How-to
Add
Review
Profile

kDavis

Logout

Discourse
Report Bugs (GitHub)
Privacy
Terms
Cookies

Add Sentences

Select Language

English (English) 

Add public domain sentences

I just made this sentence up.

Where are these public domain sentences from?

I just made this sentence up.

I confirm that these sentences are public domain and I have permission to upload them.

Submit

Sentence Collection: Review

Common Voice
moz://a

Home
How-to
Add
Review
Profile

kDavis

Logout

Discourse
Report Bugs (GitHub)
Privacy
Terms
Cookies

Review Sentences

Language to review
English (English) 

1 < 1 > 407

Low bush whipped him in the face and left no sting.

At first he puzzled over something untoward he was sure had happened.

Such a dog the wise driver kills, or turns loose.










Finish Review

Sentence Collection: Review

Common Voice
moz://a

Home How-to Add Review Profile

kDavis

Logout

Discourse Report Bugs (GitHub)
Privacy Terms Cookies

Review Sentences

Language to review
English (English)

1 < 1 > 407

Low bush whipped him in the face and left no sting.

At first he puzzled over something untoward he was sure had happened.

Such a dog the wise driver kills, or turns loose.

Finish Review

Sentence Collection: Profile

Common Voice
moz://a

Home
How-to
Add
Review
Profile

Profile: kdavis

- 1 sentences added
- 3 sentences reviewed
- ... across 1 language(s)

Your languages:

- English [remove](#)
 - 1 added by you.
 - 3 reviewed by you.

Add a language you want to contribute to

-- [Add](#)

[Logout](#)

[Discourse](#)
[Report Bugs \(GitHub\)](#)
[Privacy](#)
[Terms](#)
[Cookies](#)

4 Future of Common Voice



The background image shows a close-up aerial view of geological rock formations. The rocks exhibit distinct horizontal sedimentary layers and vertical weathering patterns, creating a complex, textured surface. A prominent, light-colored diagonal line, likely a fault or a major joint, cuts through the center of the frame. The lighting from the upper left highlights the textures and colors of the rock, ranging from deep browns to bright yellows and oranges.

4.1 Non-Corporate Partnerships

Non-Corporate Partnerships



Deutsche Gesellschaft
für Internationale
Zusammenarbeit (GIZ) GmbH



The background image shows a high-angle aerial view of a vast, rugged landscape. The terrain is dominated by reddish-brown, rocky mountains and hills, with deep, dark valleys filled with sparse vegetation. The lighting suggests either sunrise or sunset, casting long shadows and highlighting the textures of the rock faces. The overall scene has a desolate, planetary feel, reminiscent of Mars.

4.2 Corporate Partnerships

Corporate Partnerships



Рахмат!

