

moz://a

Common Voice

...e il riconoscimento vocale nella
lingua piú bella del mondo!

23/10/2021



Stefania Delprete

@astrastefania

Data Scientist in TOP-IX

Mozilla e Common Voice Rep

PyCon Italia / PyData supporter
NumFOCUS ambassador

Effective Altruism Italia
Altruismo Efficace



Agenda

- ★ **Cos'è Mozilla?**
- ★ **Common Voice**
- ★ **Vuoi contribuire?**
- ★ **CV Project**



Cos'è Mozilla?

Manifesto Mozilla

Internet è una risorsa **pubblica globale** che deve rimanere aperta e accessibile.

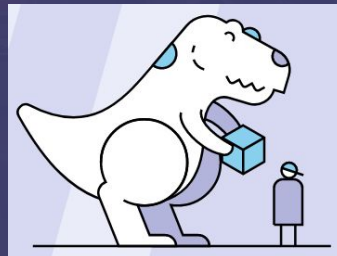
Internet deve **arricchire la vita** di ogni essere umano.

L'efficacia di Internet come risorsa pubblica dipende dal suo **carattere di interoperabilità** (relativamente a protocolli, formati di dati, contenuto), dal suo grado di innovazione e dalla partecipazione decentralizzata a livello mondiale.

La partecipazione commerciale allo sviluppo di Internet è in grado di apportare numerosi benefici, ma è fondamentale un **equilibrio tra profitto commerciale e benefici pubblici**.

La valorizzazione degli aspetti di **pubblica utilità di Internet** rappresenta un obiettivo importante, che merita tempo, attenzione e impegno.

<https://www.mozilla.org/it/about/manifesto>



Cos'è Mozilla?



Si tratta di una fondazione no profit di cui Mozilla Italia è la comunità linguistica oltreché nazionale Italiana.

moz://a

we stand for
the internet



Nella prossima
foto di gruppo
potresti esserci
anche tu!

Community portal

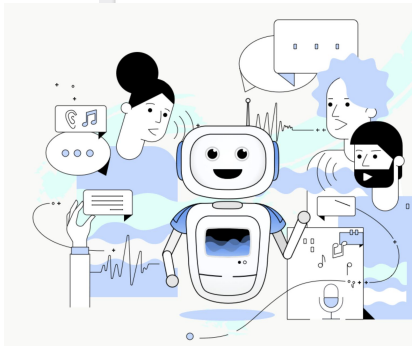


community.mozilla.org/groups/mozilla-italia

community.mozilla.org/groups/rust-italia

The screenshot shows the Mozilla Italia group page. At the top, there is a navigation bar with 'Campaigns', 'Activities', 'Events', 'Groups', and 'People'. Below this, the group name 'Mozilla Italia' is displayed in a large, bold font. Underneath the name, it says 'Verified | Italy | Created January 20, 2020'. There are three sub-navigation links: 'About us', 'Our Events', and 'Our Members'. The main content area features the Mozilla logo on the left and the text 'moz://a Italia' on the right. At the bottom right of the main content area, there is a 'Share Group' button with an upward-pointing arrow icon.

The screenshot shows the Rust Italia group page. At the top, there is a navigation bar with 'Campaigns', 'Activities', 'Events', 'Groups', and 'People'. Below this, the group name 'Rust Italia' is displayed in a large, bold font. Underneath the name, it says 'Unverified | Italy | Created May 08, 2020'. There are three sub-navigation links: 'About us', 'Our Events', and 'Our Members'. The main content area features the text 'Rust Italia' on the left and a circular logo on the right that contains a map of Italy and is surrounded by gear-like teeth.



Oct 28
Gruppo di studio di Rust
October 28, 2021 · 00:00 UTC
📍 Torino, Italy



Common Voice

Common Voice

voice.mozilla.org

Progetto di Mozilla incentrato sul raccoglimento di registrazioni vocali, per poter creare un riconoscitore vocale basato su un dataset e su un modello di pubblico dominio per tutte le lingue che partecipano al progetto.

Il modello di deep learning, basato su un paper di Baidu, è implementato dal progetto DeepSpeech di Mozilla.

<https://github.com/mozilla/voice-web>

moz://a

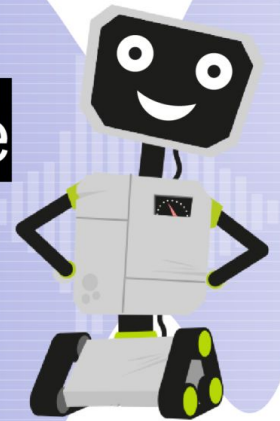
emerging

technologies



Common Voice

moz://a



Stato del progetto per la lingua inglese

- 75 879 voci
- 65 GB di dataset disponibile

- WER **11%** da v0.2.0

hacks.mozilla.org/2017/11/a-journey-to-10-word-error-rate

hacks.mozilla.org/2018/09/speech-recognition-deepspeech

Accent

23% United States English, **8%** England English,
7% India and South Asia (India, Pakistan, Sri Lanka),
3% Canadian English, **3%** Australian English,
2% Scottish English, **1%** New Zealand English,
1% Southern African (South Africa, Zimbabwe, Namibia),
1% Irish English

Age

24% 19 - 29, **13%** 30 - 39, **10%** 40 - 49, **6%** < 19,
4% 60 - 69, **4%** 50 - 59, **1%** 70 - 79

45% Male, **15%** Female

Dati al 23/10/2021

Inglese

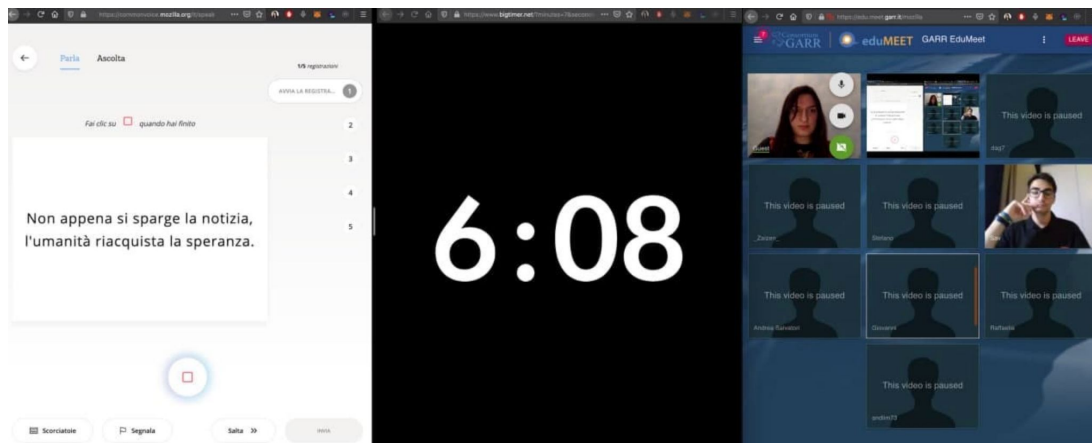
Parlato da **78,131**

Ore convalidate **2.1k / 2.5k**

CONTRIBUTE

Stato del progetto per la lingua italiana

- Oltre 1 milione di frasi da diverse fonti
- Definizione dei dialetti in corso
- Avviato il 3 Luglio 2018
- Promozione:
 - Social network
 - Linux Day
 - Italian Hacker Camp
 - MERGE-it
 - Women in Voice
 - Meetup in tutta Italia
- TU!



Stiamo

realizzando un dataset di voci multilingua e open source che chiunque può utilizzare per migliorare applicazioni munite riconoscimento vocale.

Crediamo che mettere in circolazione dataset di voci pubblici di grandi dimensioni possa promuovere innovazione e sana concorrenza commerciale tra le tecnologie di apprendimento automatico basate sul riconoscimento vocale.

Il dataset multilingue di Common Voice è già il più grande del suo genere disponibile pubblicamente, ma non è l'unico.

Considera questa pagina come un punto di riferimento per dataset open source e, man mano che Common Voice continua a crescere, una pagina per i nostri aggiornamenti.



Versione	Common Voice Corpus 7.0	▼
Lingua	Italiano	▼

DATA	2021-07-21
DIMENSIONE	8 GB
VERSIONE	it_317h_2021-07-21
TOTALE ORE CONVALIDATE	288
TOTALE ORE COMPLESSIVE	317
LICENZA	CC-0
NUMERO DI VOCI	6,407
FORMATO AUDIO	MP3
DIVISIONI	Accento Età 21% 19 - 29, 16% 30 - 39, 14% 50 - 59, 14% 40 - 49, 3% 60 - 69, 1% < 19 58% Maschio, 12% Femmina

Common Voice roadmap





Vuoi contribuire?

Come contribuire a Common Voice?

Locandina per IHC 2018

Registra la tua voce

- Leggi la frase che compare
- Scandisci tutte le parole, anche le finali!
- Leggi anche i numeri, per esteso
- **Non preoccuparti del tuo accento!**



Revisiona le registrazioni

Accertati che le registrazioni siano corrette:

- Frasi lette per intero e correttamente
- Deve essere comprensibile/udibile, quindi **assenza**, o bassi, di rumori di sottofondo!

Presta la tua **VOCE** e vinci
FANTASTICI PREMI

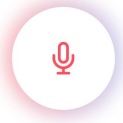
- Scarica e recensisci Firefox o Firefox Focus: **Adesivo Firefox**
- Scarica Firefox Nightly o Firefox Focus Nightly: **Spilla Firefox**
- ... e su Common Voice: **x2**
- 10 registrazioni: **Braccialetto**
- 20 registrazioni: **Pulisci-schermo**
- 30 registrazioni: **Adesivo per auto**
- 50 registrazioni: **Poster**
- 100 registrazioni: **Maglietta**
- 200 registrazioni: **Zainetto**
- 250 registrazioni: **Camicia**

—MOZILLA ITALIA—
Fino a esaurimento delle scorte. Affrettati!

1. Vai su: voice.mozilla.org/it

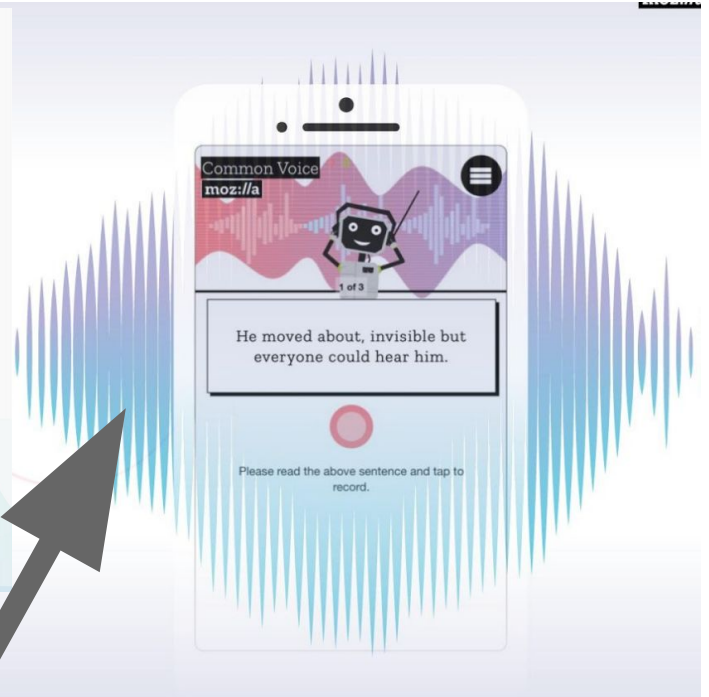
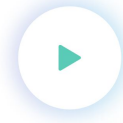
Parla

Dona la tua voce



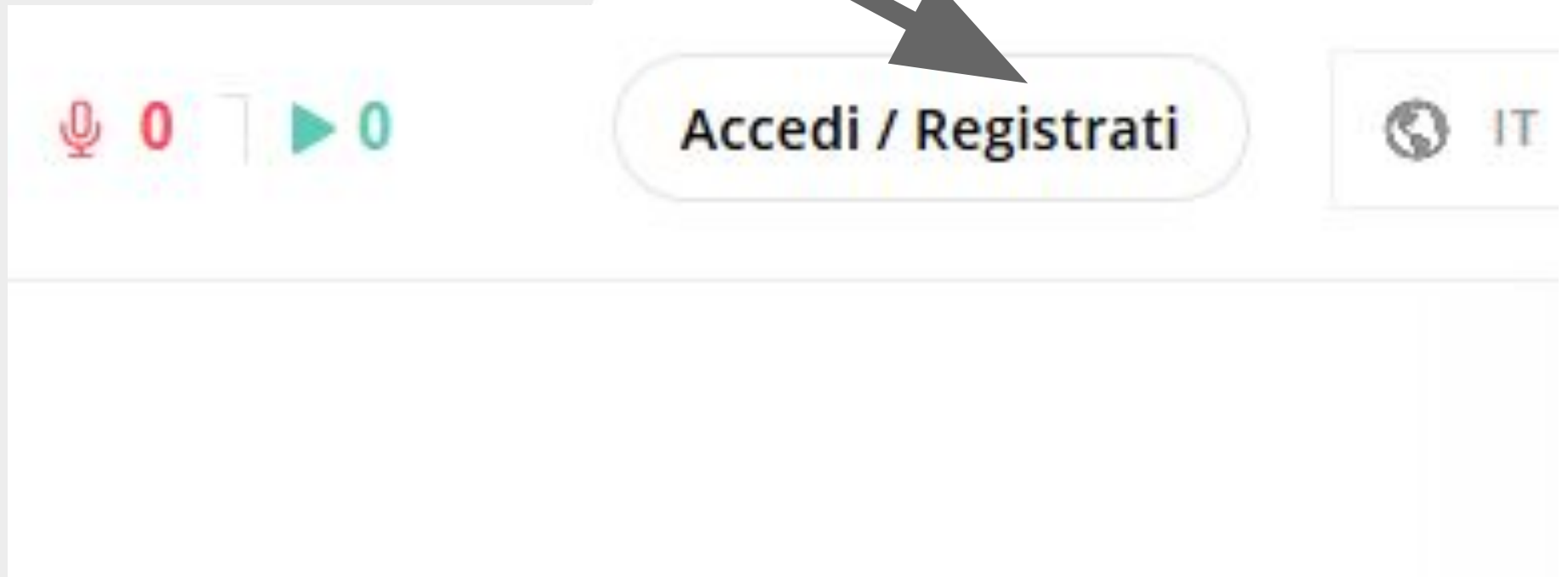
Ascolta

Aiutaci a convalidare le registrazioni



Sia da PC che da smartphone

2. Crea un account



<https://voice.mozilla.org/it>

3. Dona la tua voce!

Parla

Dona la tua voce



Ascolta

Aiutaci a convalidare le registrazioni



Obiettivo: 15 Minuti (225 registrazioni) a persona

4. Convalida le voci

Parla

Dona la tua voce



Ascolta

Aiutaci a convalidare le
registrazioni



Obiettivo: 450+ revisioni a persona

5. Controlla i tuoi progressi



[Log In / Sign Up](#)

 **225**  **450**

[Log In / Sign Up](#)

Festeggia!

 **Profilo**

 **Avatar**



 **Impostazioni**

 **Elimina profilo**

 **Scarica i miei dati**

Desideri scaricare i tuoi dati?

Scegli che cosa vuoi scaricare:

	Profilo Contiene email, nome utente e le informazioni demografiche disponibili	Qualche byte	Scarica i dati del profilo
	Registrazioni Contiene gli mp3 e le frasi correlate. Potrebbe richiedere del tempo per la preparazione	Dell'ordine dei megabyte	Richiedi le registrazioni

Richieste di download di registrazioni precedenti

Quando richiedi le tue registrazioni, vengono compilate in uno o più file ZIP. Ecco le tue richieste passate:



October 12, 2021, 6:17 PM

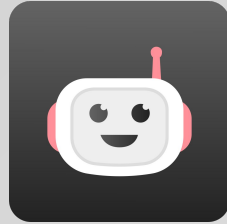
File ZIP singolo contenente 757 registrazioni vocali, dimensione totale dell'archivio 0.1 MB. Scade November 11, 2021.

 [Scarica](#)

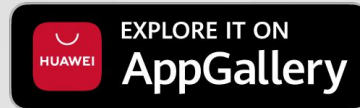
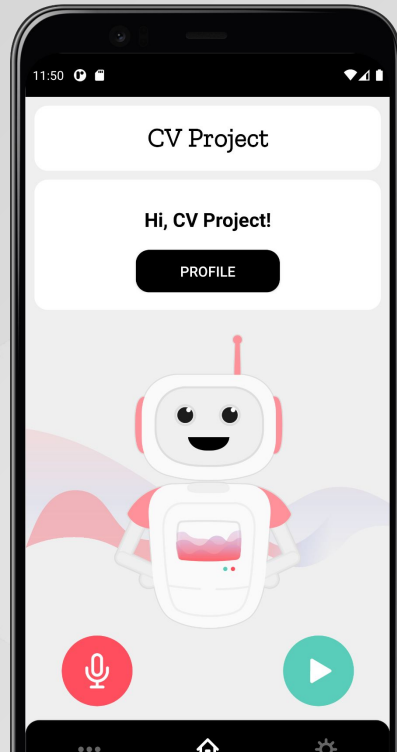
Puoi anche scaricare i tuoi dati!



CV Project



CV Project



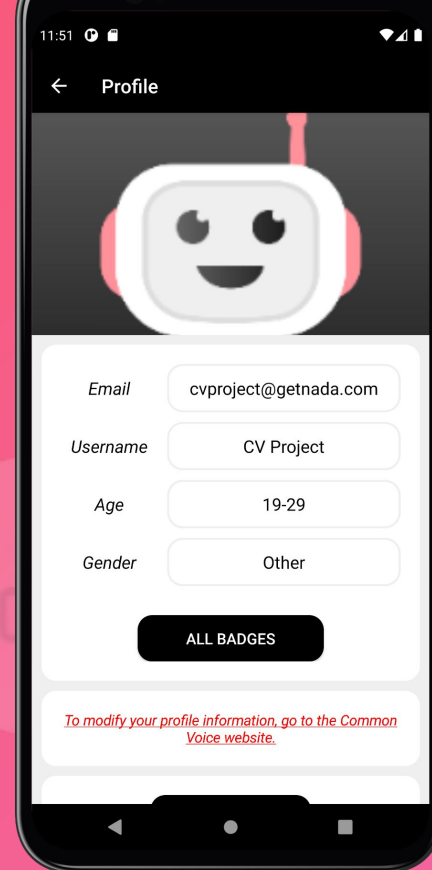
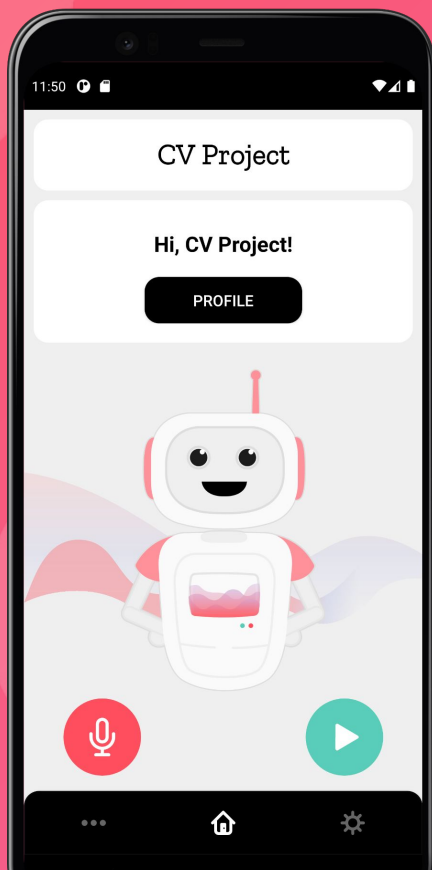
L'app **non** è sviluppata da Mozilla.
Codice sorgente disponibile su GitHub.

Perché usare l'app?

- ❑ L'app, rispetto al sito web, è **leggera** e **pulita**. Ed è **open-source**.
- ❑ **Qualità audio** degli smartphone solitamente **superiore** rispetto a quella dei portatili.
- ❑ **Personalizzazione**: tema scuro, riproduzione automatica, ecc.
- ❑ **Modalità offline**: convalida e registra anche quando non è sei connesso a Internet!
- ❑ Pensata per **tutti**: intuitiva e tutto a portata di "tap"



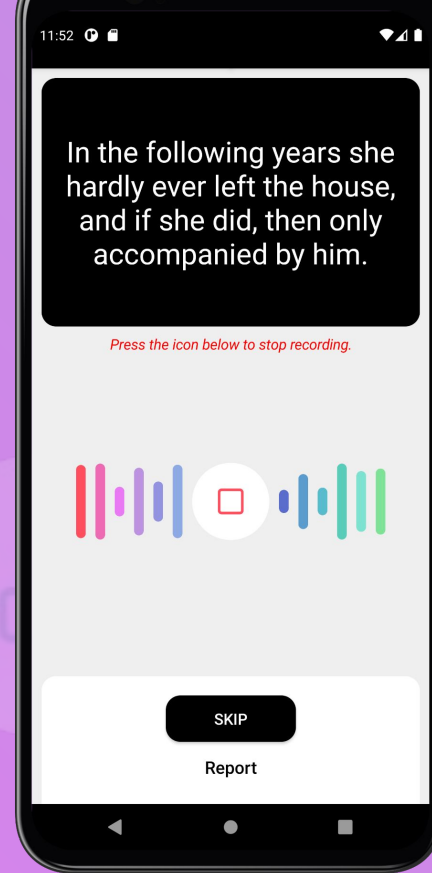
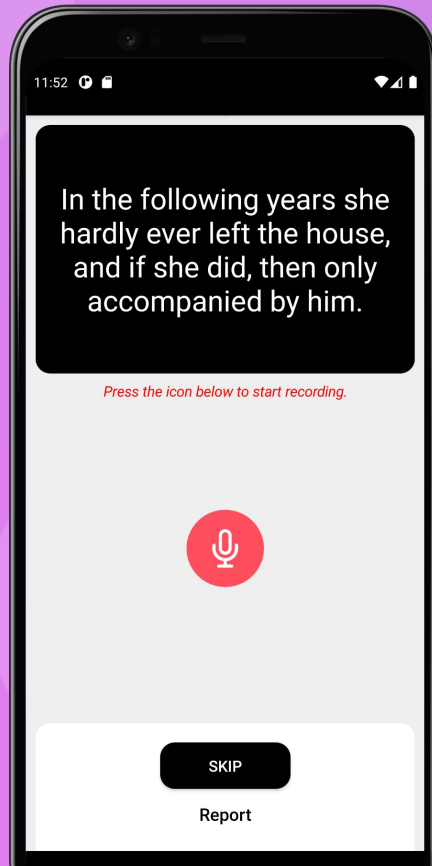
Accedi al tuo account*



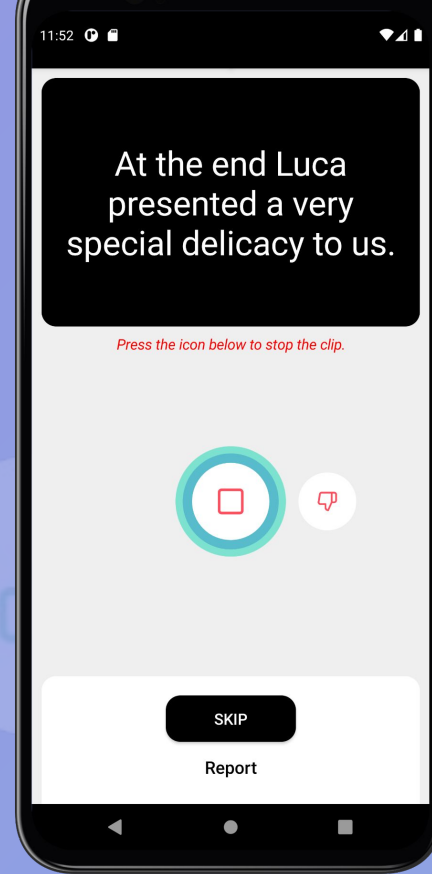
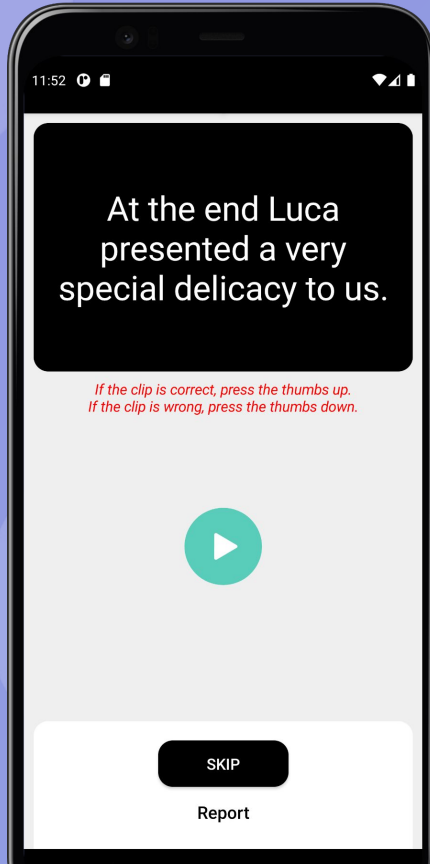
CV Project

*Puoi usare l'app anche senza effettuare l'accesso

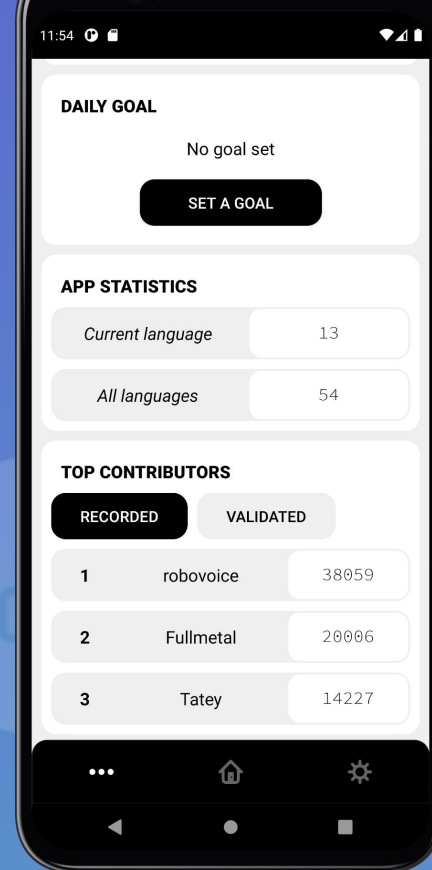
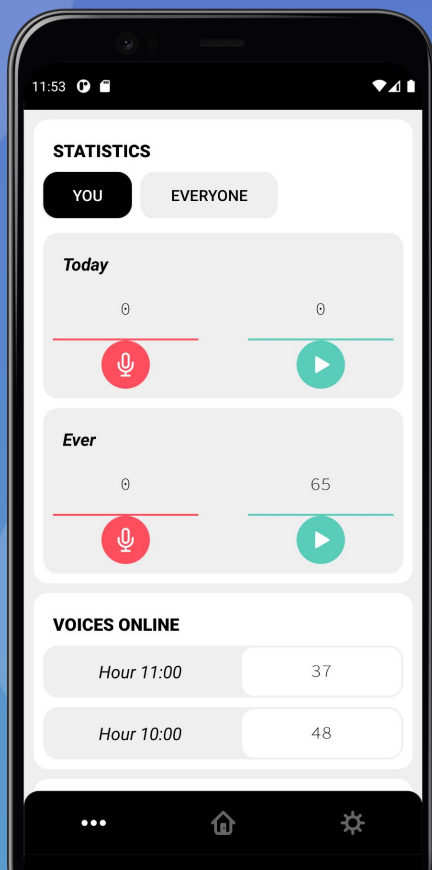
Registra frasi



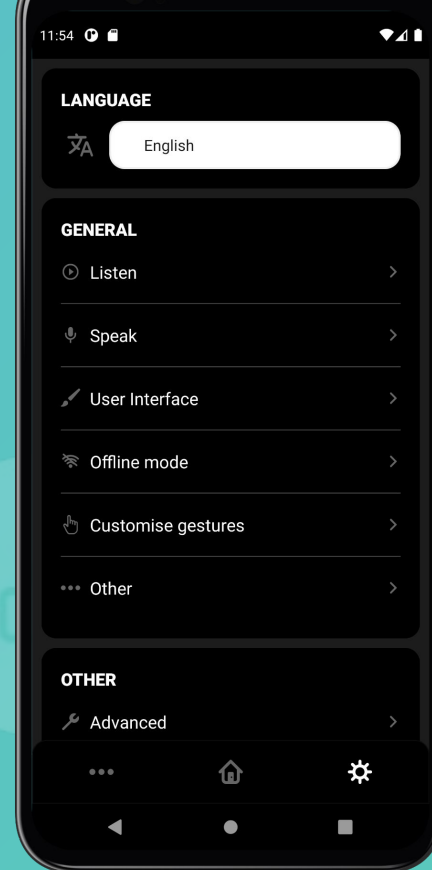
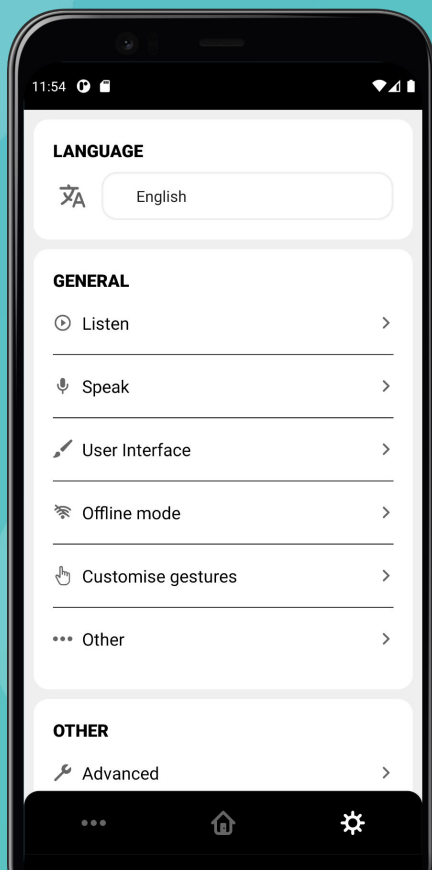
Convalida registrazioni



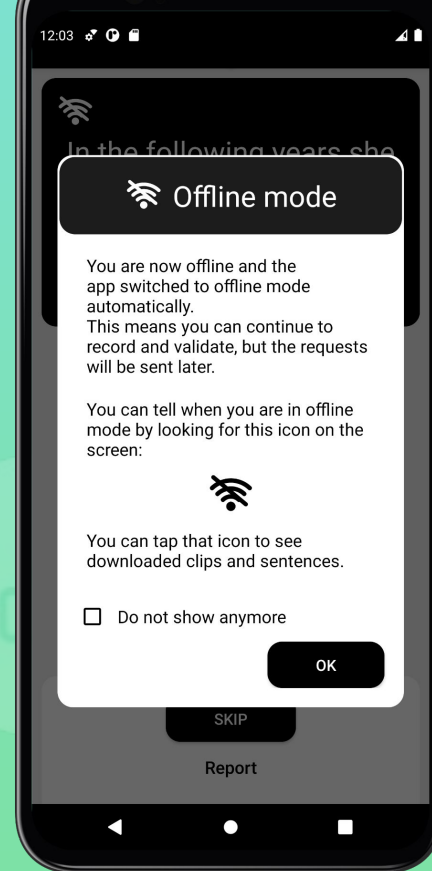
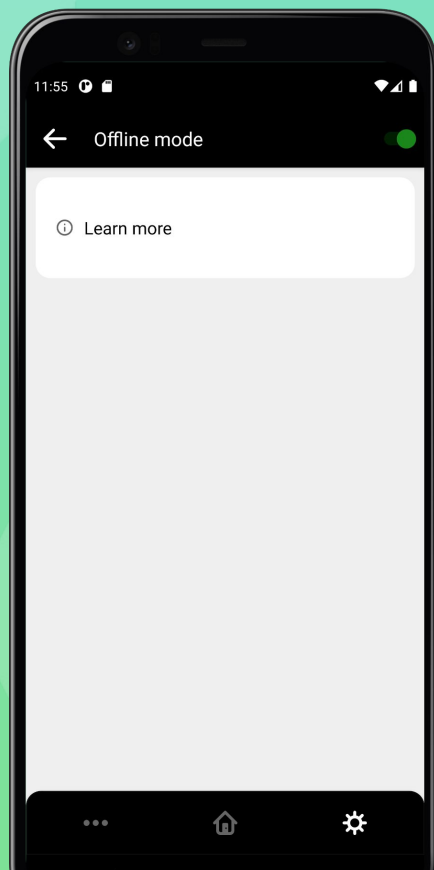
Tieni sott'occhio i tuoi progressi



Personalizza l'app secondo le tue preferenze

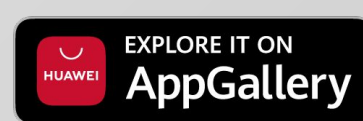


Modalità offline



Installa subito l'app!

🔍 CV Project



App disponibile anche su Amazon AppStore e GitHub.

Lascia una recensione sul tuo store preferito, e unisciti al gruppo Telegram [@common_voice_android](#).



CV Project

*L'app è sviluppata da Saverio Morelli.
Puoi scaricare l'apk anche da GitHub.*




Saluti da Stefania Delprete
@astrastefania

Restiamo in contatto!

Gruppo Telegram Italiano **@mozItabot**

Website e forum **mozillaitalia.org**

Home Contatti Download Eventi FAQ Forum







Che cos'è moz://a





Mozilla è l'organizzazione senza fini di lucro che ha realizzato Firefox, il primo browser alternativo. Da sempre promuove prodotti e linee di condotta per mantenere Internet al servizio delle persone, non dei propri profitti.

Mozilla segue i principi del suo Manifesto, principi scritti nel 1998 ma che continuano a guidarla ancora oggi. Nel 2018 è stato anche creato un addendum per mettere in rilievo **inclusione, privacy e sicurezza** per tutti in rete.

Scopri di più su Mozilla Leggi il Manifesto di Mozilla Scarica i prodotti Mozilla

Alcuni progetti Mozilla

 Firefox  Thunderbird  SeaMonkey  Common Voice

 MDN  Firefox Monitor  Firefox Relay  Facebook Container