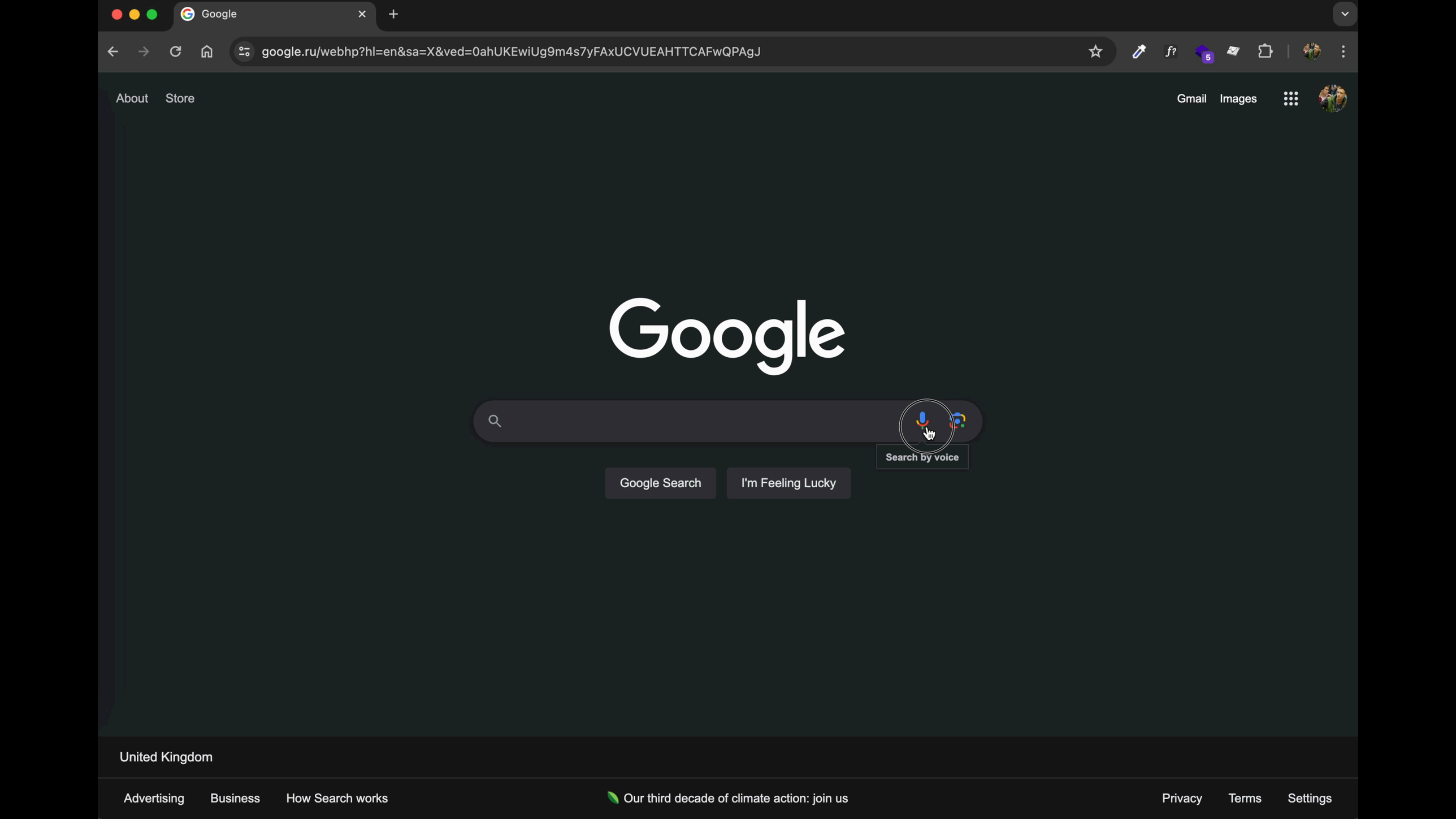
Voice Navigation

In Browser

Inspire


Google
google.ru/webhp?hl=en&sa=X&ved=0ahUKEwiUg9m4s7yFAxUCVUEAHTTCAFwQPAgJ
About
Store
Gmail
Images
Google
Search by voice
Google Search
I'm Feeling Lucky
United Kingdom
Advertising
Business
How Search works
Our third decade of climate action: join us
Privacy
Terms
Settings

Statistics of Usage

- ~50% of adults use voice search
- 3-4M smart speakers are sold in Russia every year
- Voice search shopping is expected to cross $40B

User Experience

Main Use Cases

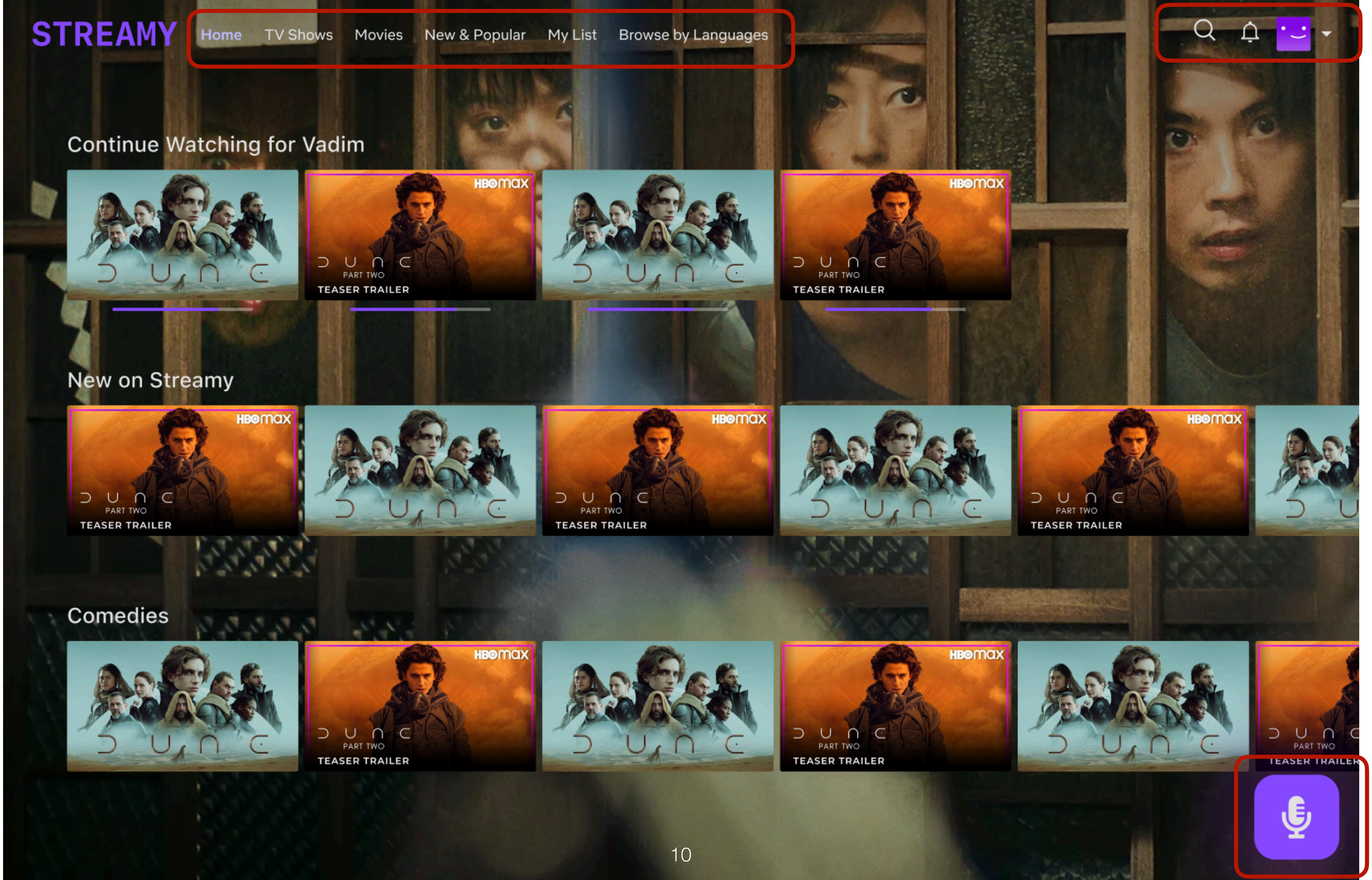
- Search
- Media content
- E-commerce

Alternative Use Cases

- Onboarding
- Accessibility
- Games

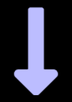
Let's build an App

with Voice Navigation

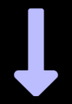

STREAMY
Home
TV Shows
Movies
New & Popular
My List
Browse by Languages
Continue Watching for Vadim
New on Streamy
Comedies
HBOmax
DUNE
PART TWO
TEASER TRAILER

How it works?

Recognising Speech



Get Route by Phrase



Navigate

Step 1.
Recognising Speech

Implementations

- Web Speech API
- Custom ML Models in Browser
- Speech Recognition at Server

Web Speech API

Voice in Browser

2011

Draft
Specification

2012

Specification
Published

2013

First
Release

Native API Overview

Speech API Components

- Speech Synthesis
- Speech Recognition
- Speech Grammar

Speech Synthesis

```
const utterThis = new SpeechSynthesisUtterance('Привет, Holy JS')
const voices = speechSynthesis.getVoices()
const ruVoice = voices.find(voice => voice.lang === 'ru-RU')
if (ruVoice) utterThis.voice = ruVoice
speechSynthesis.speak(utterThis);
```

Browser Compatibility

Speech Synthesis API - UNOFF

A web API for controlling a text-to-speech output.

Usage: % of all users — Global 94.62%

Current aligned | Usage relative | Date relative | Filtered | All

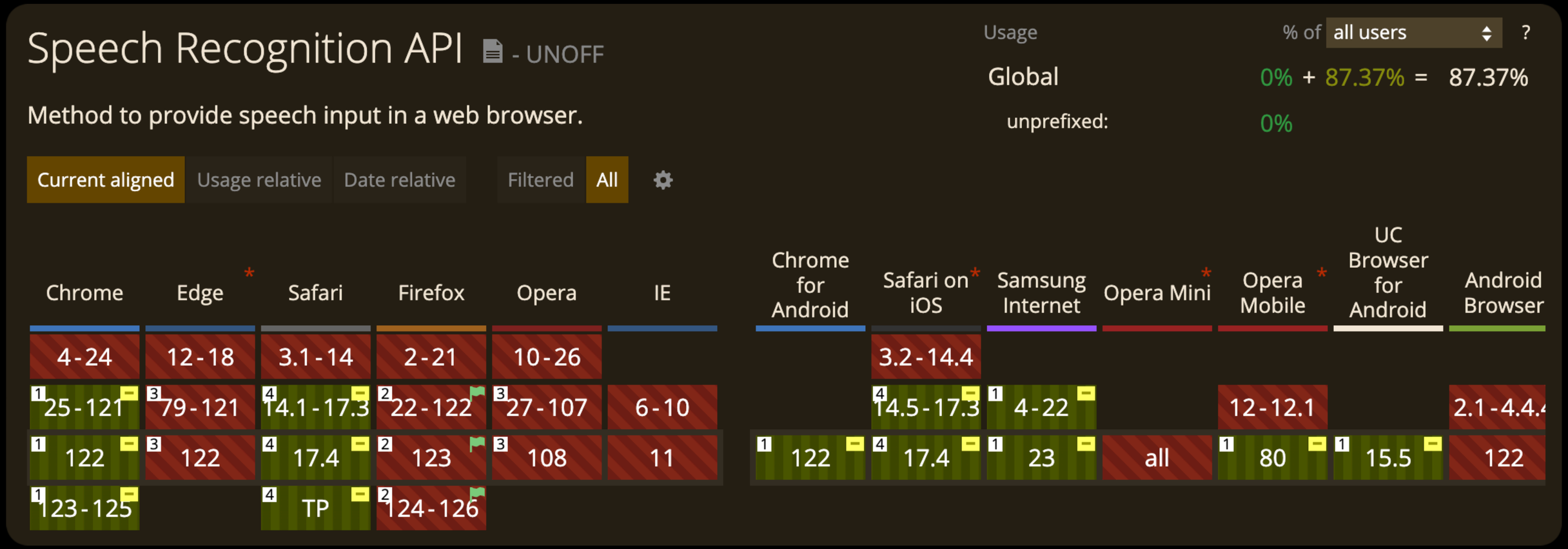
Chrome	Edge *	Safari	Firefox	Opera	IE
4-32	12-13		2-30	10-26	
33-54	14-18	3.1-6.1	1 31-48	27-63	
2 55-121	2 79-121	7-17.3	49-122	2 64-107	6-10
2 122	2 122	17.4	123	2 108	11
2 123-125		TP	124-126		

Chrome for Android	Safari on iOS *	Samsung Internet	Opera Mini *	Opera Mobile *	UC Browser for Android	Android Browser
	3.2-6.1	4				
	7-17.3	5-22		12-12.1		2.1-4.4.4
122	17.4	23	all	80	15.5	122

Speech Recognition

```
> const recognition = new webkitSpeechRecognition();
  recognition.lang = "ru-RU";
  recognition.addEventListener("result", (e) => {
    const transcript = Array.from(e.results).map(result => result[0].transcript).join('');
    console.log(transcript);
  });
  recognition.start();
```

Browser Compatibility



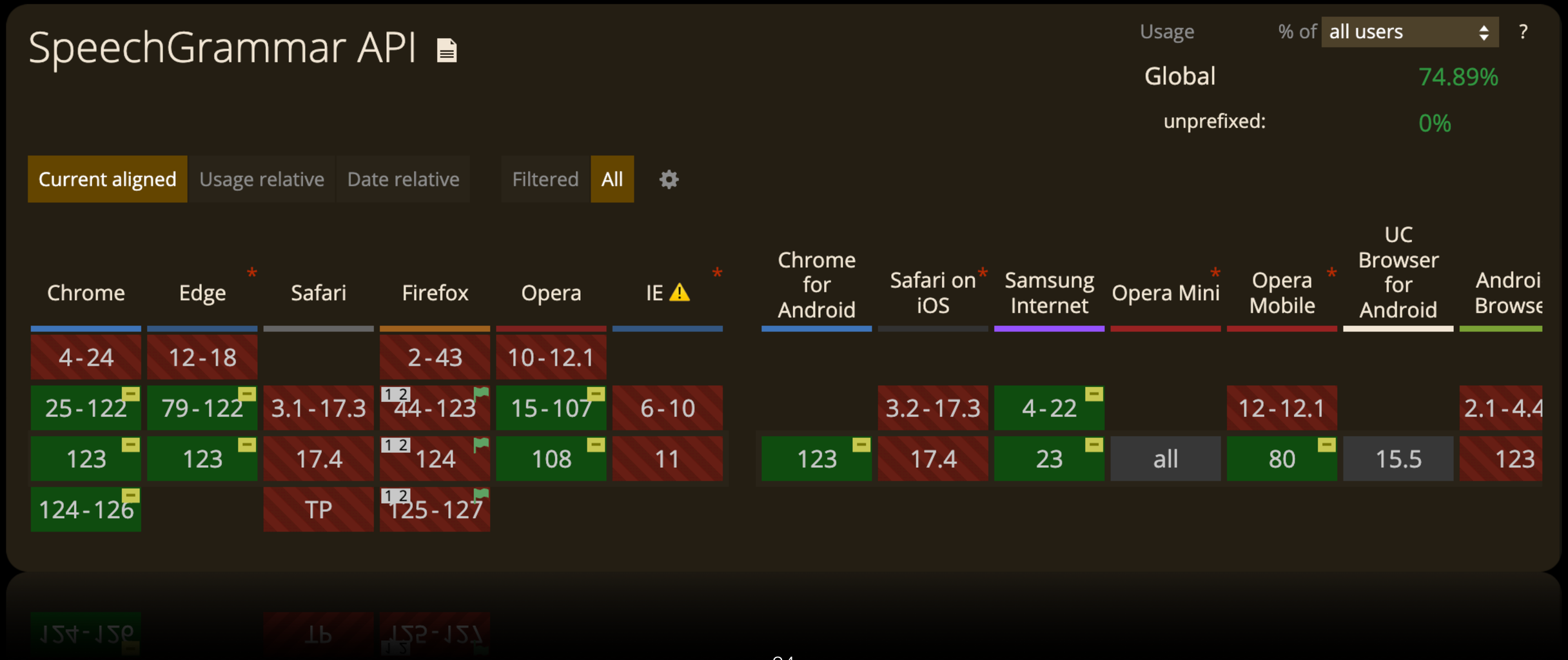
Speech Grammar

```
107   const grammar =
108     "#JSGF V1.0; grammar colors; public <color> = aqua | azure |
109   const speechRecognitionList = new SpeechGrammarList();
110   speechRecognitionList.addFromString(grammar, 1);
111   speechRecognition.grammars = speechRecognitionList;
```

Speech Grammar

```
1    public <command> = <action> <object>;
2    <action> = [please] (open | close);
3    <object> = (window | door | application);
```

Browser Compatibility



Other Devices

- Mobile
- TV

Setup

IDE Support

@types/dom-speech-recognition

Recognise phrase in Browser

```
90  const createSpeechRecognition = ({
91    lang,
92    onResult,
93  }: ICreateSpeechRecognition) => {
94    window.SpeechRecognition = window.SpeechRecognition || window.webkitSpeechRecognition
95    const CONFIDENCE_RATE = 0.8
96
97    if (!SpeechRecognition) {
98      loglevel.error('Speech recognition API is not supported in this browser.')
99      return
100   }
101
102   const speechRecognition = new SpeechRecognition()
103   speechRecognition.continuous = false
104   speechRecognition.interimResults = false
105   speechRecognition.lang = lang
```

Polyfill Problem

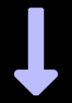
Polyfill types

- Custom ML Model in Browser
- Speech Recognition at Server

Custom ML Model in Browser

How it works?

Download ML Model



Run Model

Record audio

Recognise text from audio

HOLY JS /...

- Спикер

Алексей Охрименко

Яндекс

- Эксперт

Зар Захаров

ВКонтакте

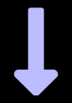
TL;DR

- High recognition quality
- Support multi-languages
- ML size is huge for Browser

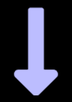
Speech Recognition at Server

How it works?

Setup socket connection

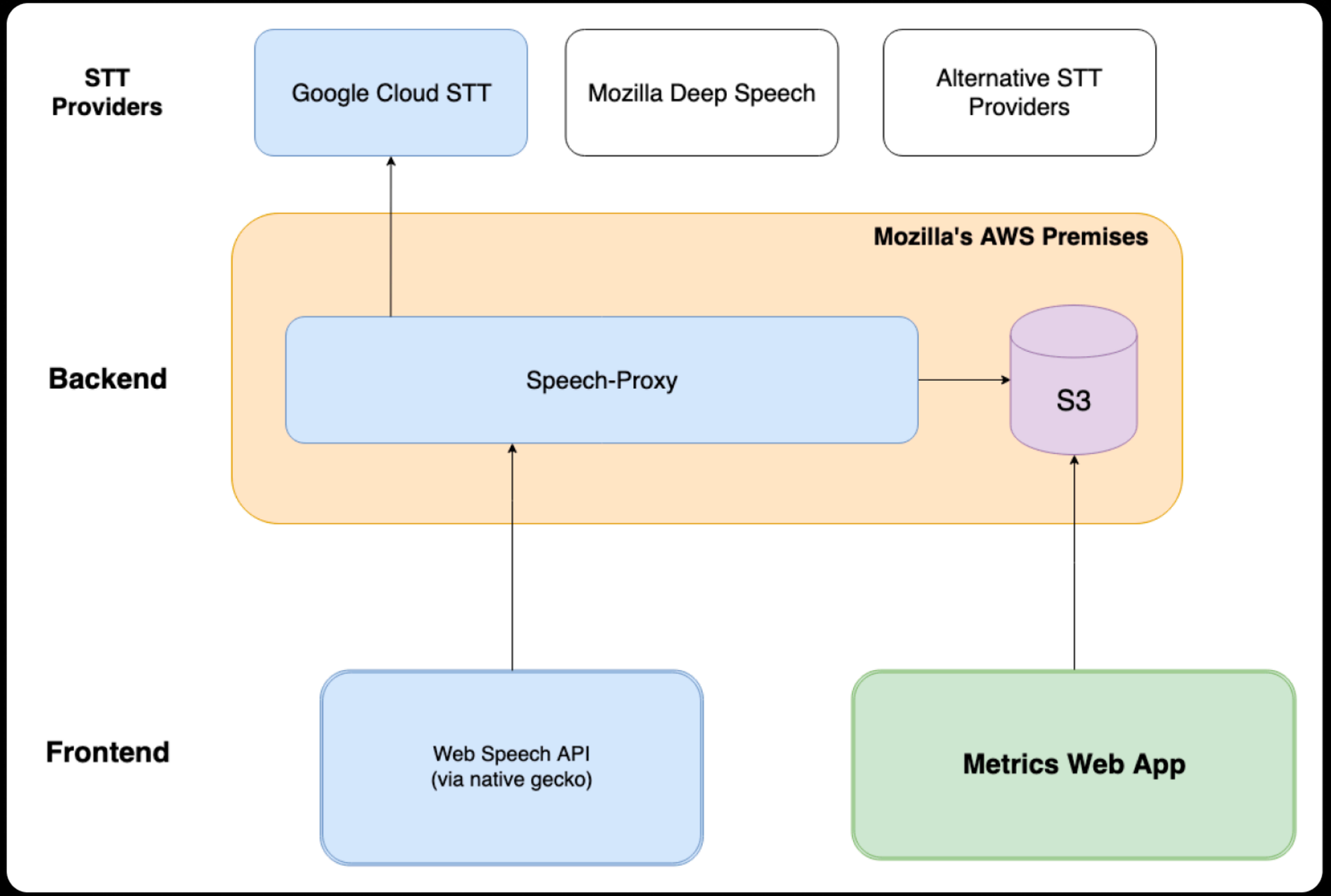


Get audio stream and record



Send record to server

Server recognises and returns phrase


STT Providers
Google Cloud STT
Mozilla Deep Speech
Alternative STT Providers
Mozilla's AWS Premises
Backend
Speech-Proxy
S3
Frontend
Web Speech API
(via native gecko)
Metrics Web App

Available Server Polyfills

- Speechly
- Serpia
- Azure Cognitive
- AWS Transcribe

Providers with no Polyfill

- Google Speech-to-Text
- Yandex SpeechKit
- OpenAI Speech-to-Text

Pitfalls

Multi-language Recognition problem

Headphones

On Device Change

Step 2.
Get Route By Phrase

Routes List

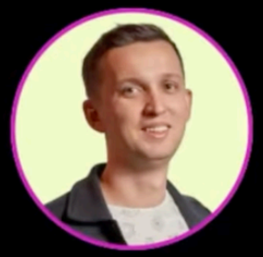
```
29  export const routes: Route[] = [
30    {
31      path: '/',
32      alias: ['Go home', 'Home page', 'Main page', 'Start page'],
33      element: <Home />,
34    },
35    {
36      path: '/my-list',
37      alias: ['Go to my list', 'My watchlist', 'My favorites', 'Show my list'],
38      element: <MyListScreen />,
39    },
40    {
41      path: '/search',
42      alias: ["Go to search", "Open search", 'Search :term', 'Find :term'],
43      element: <SearchScreen />,
44    },
```

Convert phrase into route

```
144  const getRouteByPhrase = (routes: Route[], userPhrase: string) => {
145    const routesWithWeights = routes.map(route => {
146      const aliasWithWeights = route.alias.map(routePhrase => levinstein(userPhrase, routePhrase))
147      return ({
148        ... route,
149        weight: Math.min( ... aliasWithWeights)
150      })
151    })
152    const sorted = routesWithWeights.sort((a, b) => a.weight - b.weight)
153    const matchedRoute = sorted?.[0].path
154    loglevel.debug({ userPhrase, matchedRoute })
155    return matchedRoute
156  }
```

HOLY JS /...

● Спикер



Никита Воробьёв

СберТех

● Эксперт

Ярослав Лосев

Яндекс Путешествия

Pitfalls

Args parsing problem

Routes List

```
29  export const routes: Route[] = [
30    {
31      path: '/',
32      alias: ['Go home', 'Home page', 'Main page', 'Start page'],
33      element: <Home />,
34    },
35    {
36      path: '/my-list',
37      alias: ['Go to my list', 'My watchlist', 'My favorites', 'Show my list'],
38      element: <MyListScreen />,
39    },
40    {
41      path: '/search',
42      alias: ["Go to search", "Open search", 'Search :term', 'Find :term'],
43      element: <SearchScreen />,
44    },
```

Step 3.
Navigate

Navigate user

```
159  export const AppRouter: FC = () => {
160    const navigate = useNavigate()
161    const speechRecognitionRef = useRef(createSpeechRecognition({
162      lang: 'en-US',
163      onResult(phrase) {
164        const route = getRouteByPhrase(routes, phrase)
165        navigate(route)
166      }
167    }))
```

Demo

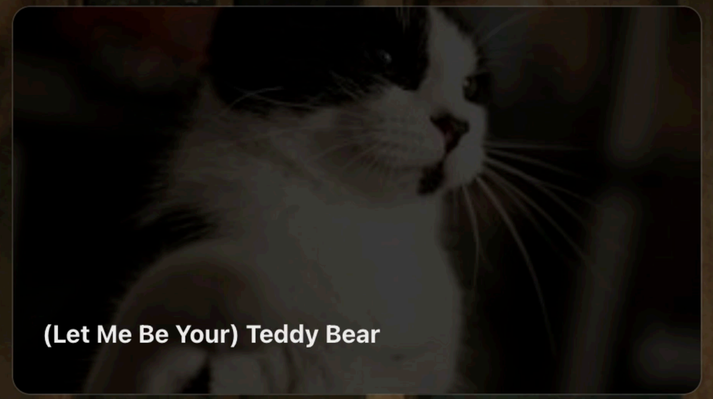
StreamyFrontend

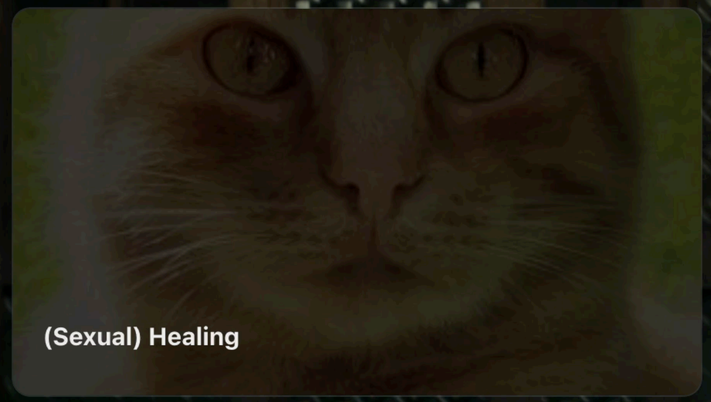
localhost:4200/categories/comedies

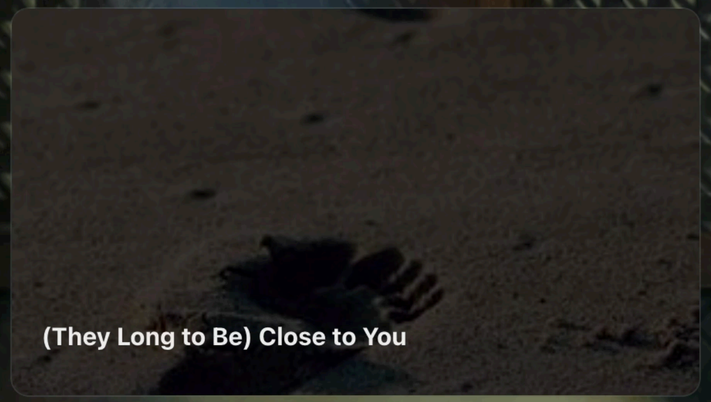
STREAMY

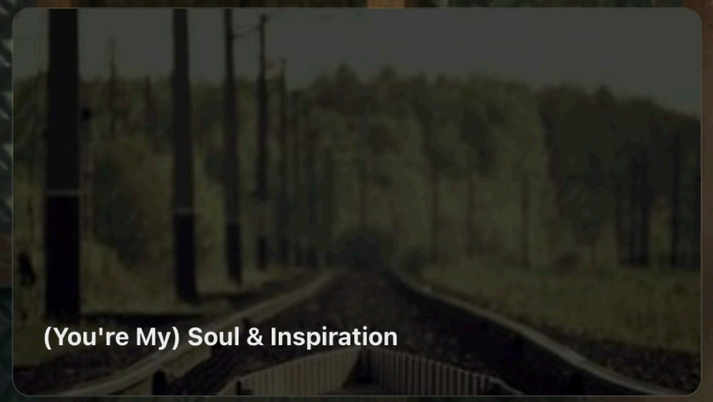
Home New & Popular Series Dramas Animations My List

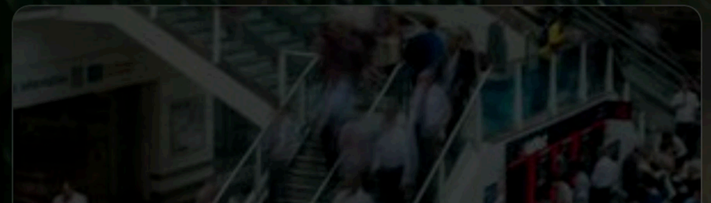
Comedy

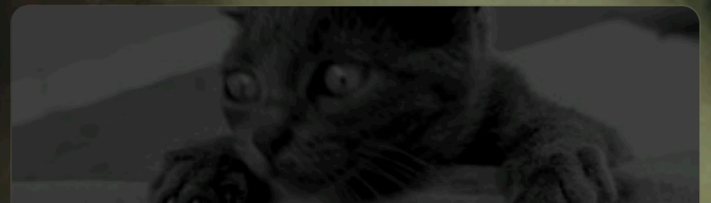


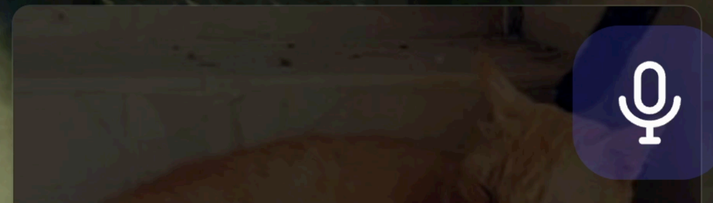












Repository

github.com/mdusafp/voice-navigation-in-browser

Experimental

Speakable schema.org

```
<script type="application/ld+json">
{
 "@context": "https://schema.org/",
 "@type": "WebPage",
 "name": "Quick Brown Fox",
 "speakable":
 {
  "@type": "SpeakableSpecification",
  "xPath": [
    "/html/head/title",
    "/html/head/meta[@name='description']/@content"
    ]
  },
 "url": "https://www.example.com/quick-brown-fox"
 }
</script>
```

Q&A

Vadim Maliutin

Telegram

LinkedIn