

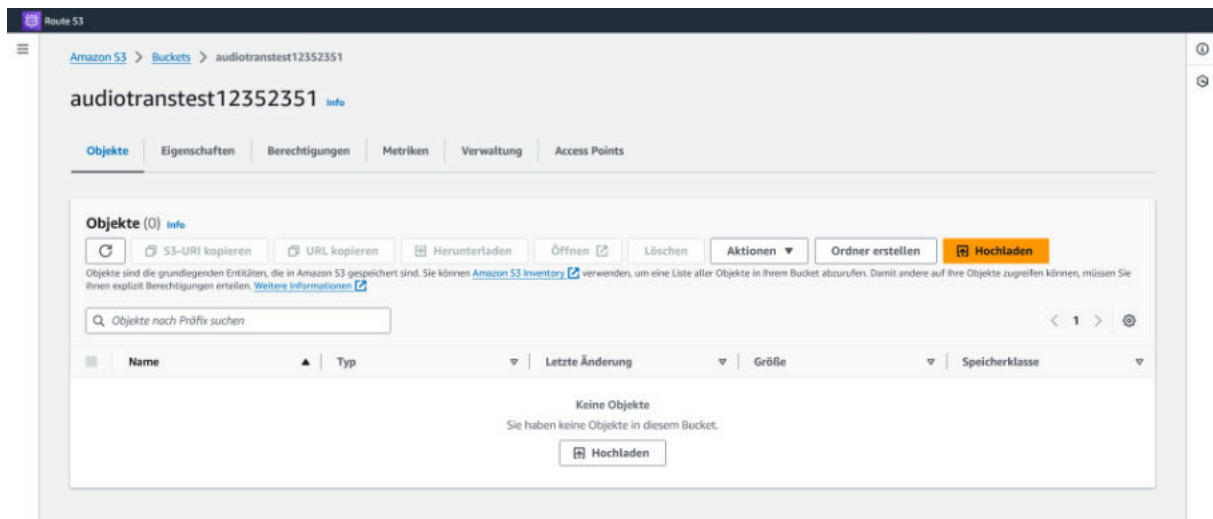
Audio Transcription mit **AWS Transcribe** und **Python / AI**



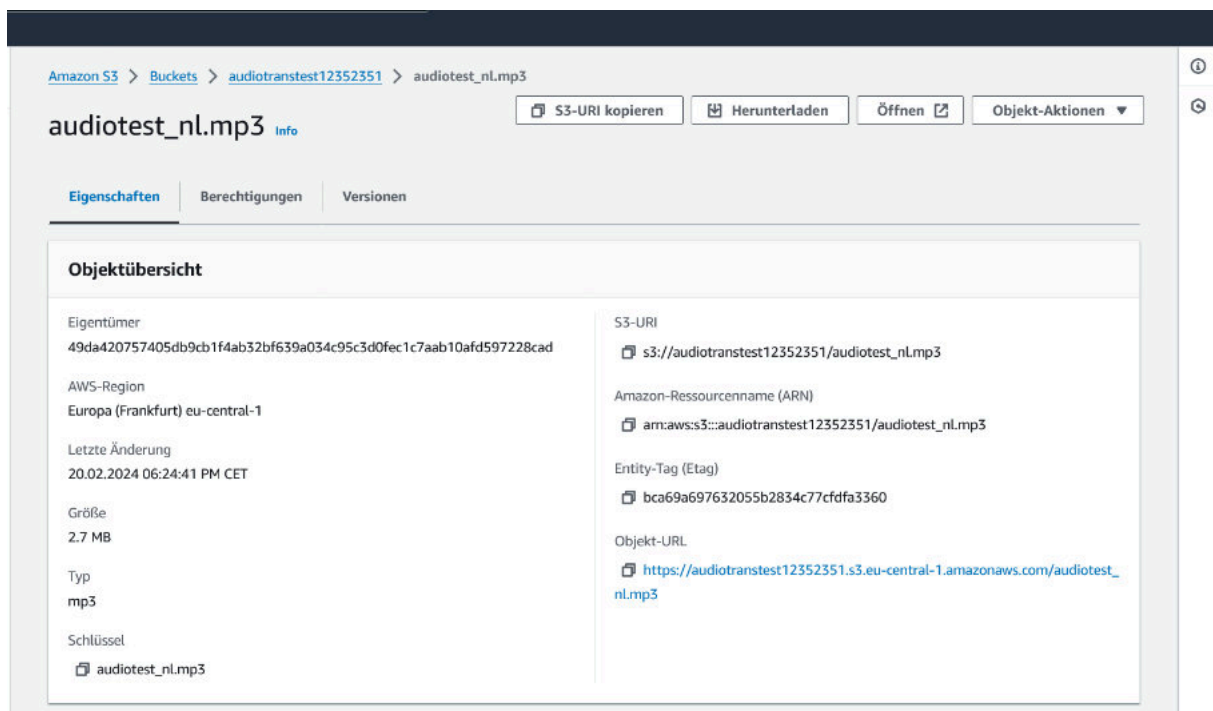
Kapitel 1 - Transcribing mit AWS - S3 Bucket erstellen

Um den Audio Transcription Service von AWS zu nutzen, müssen wir einen S3 Bucket erstellen. Ich habe den Bucket sehr offen erstellt, um etwas Zeit zu sparen (shame on me), da es nur um eine Demo geht, und mein Audiomaterial ein Rap-Song aus den Niederlanden ist, bei dem ich online keine Lyrics gefunden habe.

Wenn man „echtes“ Audiomaterial transkribieren möchte, sollte man strikt auf IAM Permissions und eventuelle Copyright-Bedingungen achten.



Nach der Erstellung des Bucket haben wir eine mp3 Datei hochgeladen.



Kapitel 1 - Transcribing mit AWS - Auftrag anlegen

Als nächstes müssen wir einen Auftrag anlegen. Das Interface von AWS Transcribe ist, im Vergleich zu anderen Services, sehr einfach und selbsterklärend.

Wir kopieren die URL unseres S3 Buckets in den Auftrag und geben wahlweise als Ziel ebenfalls einen Bucket / Ordner an, oder belassen es bei den Standard-Einstellungen, die die Ausgabe im AWS Interface anzeigt. Ich empfehle hierbei, ebenfalls einen Bucket zu wählen, das ist übersichtlicher.

Route 53

Amazon Transcribe

Transkriptionsaufträge > Transkriptionsauftrag erstellen

Step 1
Auftragsdetails angeben

Step 2
Auftrag konfigurieren - optional

Auftragsdetails angeben

Auftragseinstellungen

Name

Test

Der Name kann bis zu 200 Zeichen lang sein. Gültige Zeichen sind a-z, A-Z, 0-9, _ (Unterstrich) und - (Bindestrich).

Spracheinstellungen

Sie können Ihre Audiodatei in eine von Ihnen angegebene Sprache transkribieren oder sie durch Amazon Transcribe identifizieren und in die vorherrschende Sprache transkribieren lassen.

☒ **Bestimmte Sprache** [Informationen](#)

Wenn Sie die Sprache kennen, die in Ihrem Quellaudio gesprochen wird, wählen Sie diese Option, um die genauesten Ergebnisse zu erhalten. Die verfügbaren Optionen für die zusätzliche Verarbeitung variieren je nach Sprache.

☐ **Automatische Sprachidentifikation** [Informationen](#)

Wenn Sie die Sprache, die in Ihren Audiodateien gesprochen wird, nicht kennen, wählen Sie diese Option. Sie haben Zugriff auf weniger Optionen für die zusätzliche Verarbeitung, als wenn Sie **Bestimmte Sprache** wählen.

☐ **Automatische Identifizierung mehrerer Sprachen** [Informationen](#)

Wenn in Ihren Audiodateien mehrere Sprachen gesprochen werden und Sie sich nicht sicher sind, um welche Sprachen es sich handelt, wählen Sie diese Option. Diese Auswahl bietet im Vergleich zu **Bestimmte Sprache** begrenzte zusätzliche Verarbeitungsoptionen.

Sprache

Wählen Sie die Sprache des eingehenden Audios.

Niederländisch, NL (nl-NL)

Modelltyp [Informationen](#)

Wählen Sie den Modelltyp aus, der für den Transkriptionsauftrag verwendet werden soll.

☒ **Allgemeines Modell**

Wählen Sie diese Option, um ein Modell zu verwenden, das nicht auf einen bestimmten Anwendungsfall spezialisiert ist. Die Konfigurationsoptionen variieren je nach Sprache.

☐ **Benutzerdefiniertes Sprachmodell**

Wählen Sie diese Option, um ein Modell zu verwenden, das Sie für Ihren spezifischen Anwendungsfall trainiert haben. Dieses Modell hat weniger Konfigurationsoptionen als das allgemeine Modell.

Zusätzliche Einstellungen

Auftragswarteschlange - optional [Informationen](#)

Ermöglicht es Ihnen, Aufträge einzureichen, die das Limit für gleichzeitige Aufträge überschreiten (100). Sie müssen Zugriffsberechtigungen für die Ressourcen angeben, die eine Auftragswarteschlange verwendet.

☐ **Zur Auftragswarteschlange hinzufügen**

Eingabedaten

Speicherort der Datei in S3 eingeben

Geben Sie eine Audio- oder Videoeingabedatei in Amazon S3 an.

S3 URI

s3://bucket/prefix/file.extension

[View](#) [Browse S3](#)

Format: s3://bucket/prefix/file.extension.

Gültige Dateiformate: MP3, MP4, WAV, FLAC, AMR, OGG und WebM.

Ausgabedaten

Informationen zum Speicherort der Ausgabedaten [Informationen](#)

☐ **Serviceverwalteter S3-Bucket**

Die Ausgabe wird nach 90 Tagen entfernt, wenn der Auftrag abläuft.

☒ **Vom Kunden angegebener S3-Bucket**

Kapitel 1 - Transcribing mit AWS - Auftrag anlegen

Auf der nächsten Seite des Auftrags können wir ein paar Sonderoptionen festlegen, zum Beispiel, wie viele Alternative Erkennungen wir behalten möchten.

Die Alternativerkennungen sind nach unserem Test unverzichtbar, machen es allerdings im Nachgang schwieriger, die JSON-Ausgabe wieder zusammen zu puzzeln - man hat mehrere Alternativversionen eines Wortes, z.B. „ein“ - „eine“ - „eines“ in einem JSON-Objekt. Das wird schnell unübersichtlich.

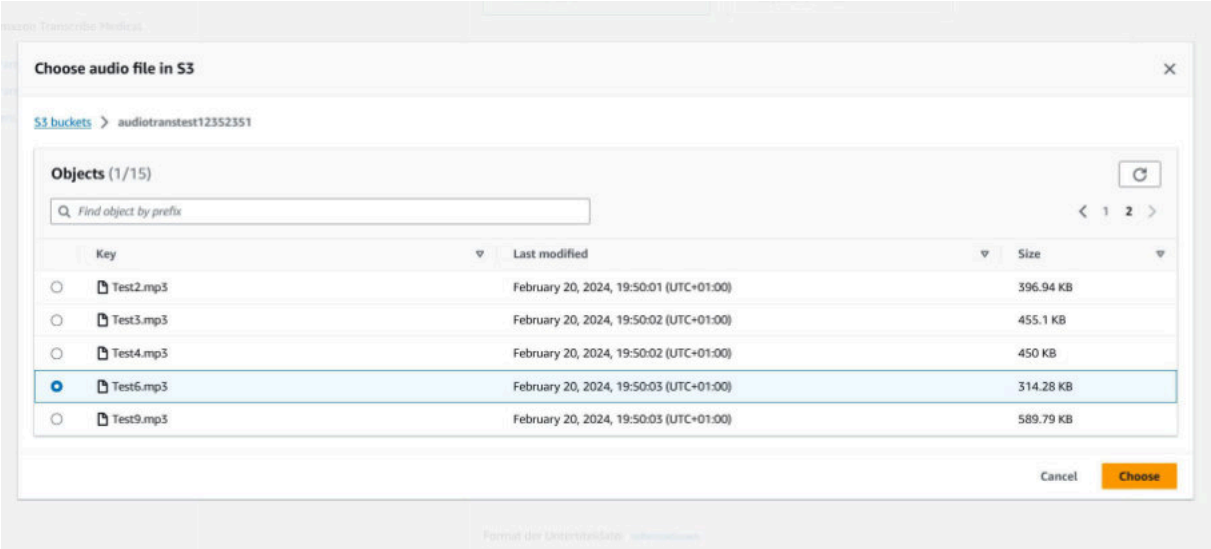
The screenshot shows the Amazon Transcribe console interface. On the left is a navigation menu with options like 'Transkription in Echtzeit', 'Transkriptionsaufträge', and 'Anruf-Analytics von Amazon Transcribe'. The main area is titled 'Auftrag konfigurieren - optional' and contains several sections for configuring transcription options:

- Audioeinstellungen:** Includes 'Audioidentifikation' (selected), 'Alternative Ergebnisse' (selected), and a field for 'Maximale Anzahl an Alternativen' set to 3.
- Inhaltsentfernung:** Includes 'PII-Redigierung' (selected) and 'Wortschatzfilterung'.
- Personalisierung:** Includes 'Benutzerdefinierter Wortschatz'.

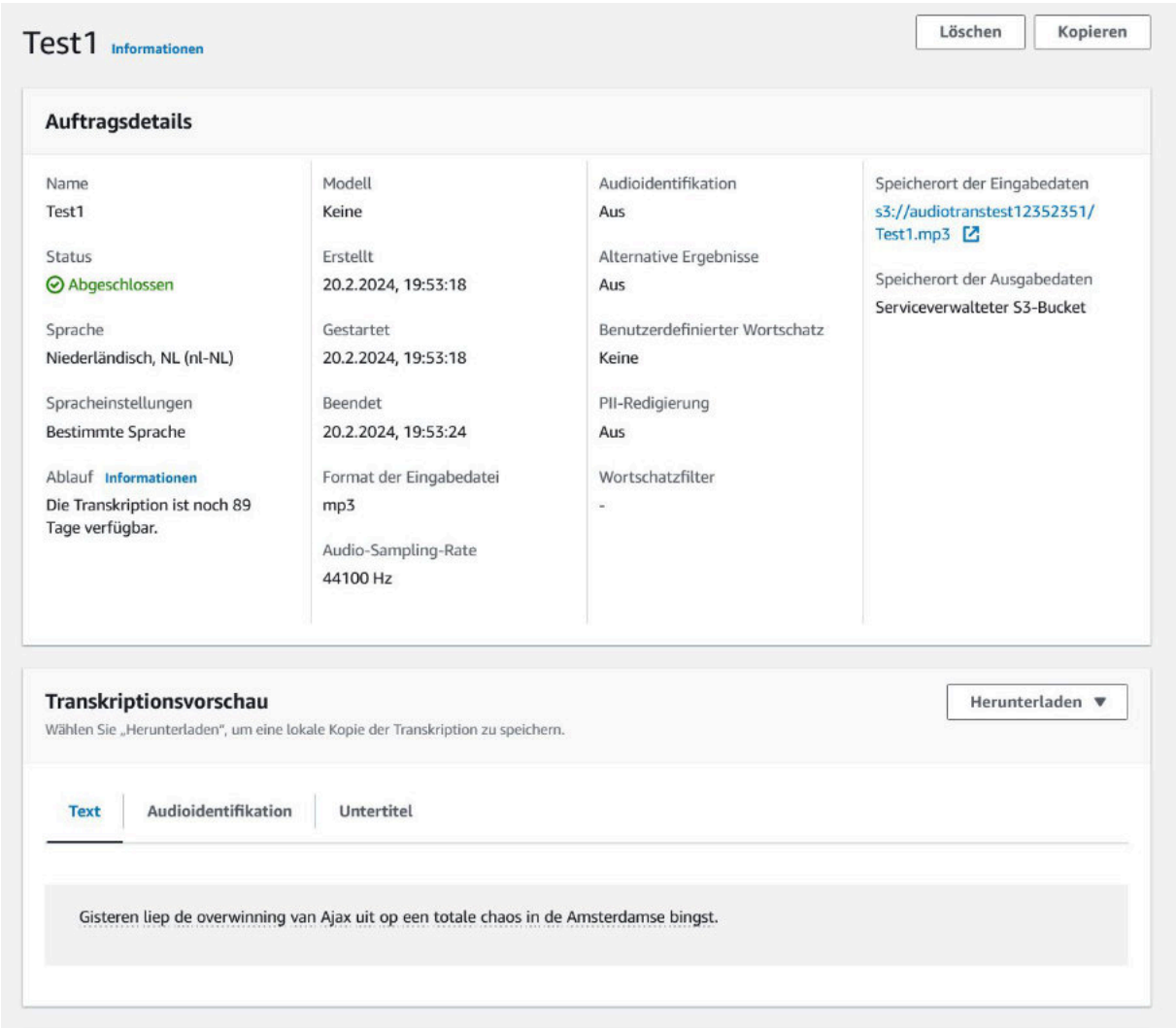
At the bottom right, there are buttons for 'Stornieren', 'Zurück', and 'Auftrag erstellen'.

Kapitel 1 - Transcribing mit AWS - Auftrag anlegen

In der Übersicht sieht man die Aufträge. Da es beim ersten mal nicht wirklich gut funktioniert hatte, haben wir das Audio mit professioneller Software nachbearbeitet und die Stimme herausgeholt. Die Software RX 10 Advanced von iZotope hat in der Anschaffung über 1000€ gekostet und sehr gute Arbeit geleistet. Mehr dazu im nächsten Kapitel.



Der erste Auftrag ist fertig.



Kapitel 1 - Transcribing mit AWS - Auftrag auswerten

Die weiteren Aufträge werden ebenfalls schnell fertig.

Der Transkriptionsauftrag wurde erfolgreich erstellt.

Amazon Transcribe > Transkriptionsaufträge

Transkriptionsaufträge (15) Informationen

Herunterladen Kopieren Löschen Auftrag erstellen

Q Auftragsnamen suchen Status: Alle

Name	Status	Sprache	Spracheinstellungen	Modelltyp	Modellname	Erstellt
Test14	In Bearbeitung	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:55
Test13	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:55
Test12	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:58
Test11	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:58
Test10	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:58
Test9	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:57
Test8	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:57
Test7	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:56
Test6	Abgeschlossen	Englisch, US (en-US)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:56
Test5	Abgeschlossen	Niederländisch, NL (nl-NL)	Bestimmte Sprache	Allgemein	-	February 20 2024, 19:55

Wir nutzen Bash Tools, um den Text aus dem etwas unübersichtlichen JSON Objekt zu extrahieren. Hier sieht man ebenfalls die Alternativ-Versionen.

```
$ cat Test.json | jq '.. | .transcript? | select(. != null)'
"Gisteren liep de overwinning van Ajax uit op een totale chaos in de Amsterdamse binnenstad. al van en van clubs zijn. We gaan je staat man in het zwart, petje op zak, spot is op pad en ik wacht op contact. tas. geen staande. concentraat. we lopen. de zit. kan een shirt heel snel uitdoen of wegdoen. trap echt je koptrap... en anders ga ik jullie dat hele kroegje aan kan. twee onze vak is van clubs. een Ginkel van glas stormmap. stille stap Mensen vechten nog binnen. Zoveel actie, het lijkt net een film. Jij bent met sukkeels en ik ben
```

```
$ cat Test.json | jq | grep -i transc
"transcripts": [
  "transcript": "Gisteren liep de overwinning van Ajax uit op een totale chaos in de Amsterdamse binnenstad. al van en van clubs zijn. We gaan je staat man in het zwart, petje op zak, spot is op pad en ik wacht op contact. tas. geen staande. concentraat. we lopen. de zit. kan een shirt heel snel uitdoen of wegdoen. trap echt je koptrap... en anders ga ik jullie dat hele kroegje aan kan. twee onze vak is van clubs. een Ginkel van glas stormmap. stille stap Mensen vechten nog binnen. Zoveel actie, het lijkt net een film. Jij bent met sukkeels en ik ben met killers. op richting de vakten er rellen van clubs een ik van Xanders heb gezien en niet hierzo. Vooral die mensen had ik echt meteen je kopterk getrokken scanner. I saw that there is some terrorist attack. # R",
  "transcript": "Gisteren liep de overwinning van Ajax uit op een totale chaos in de Amsterdamse binnenstad.",
  "transcript": "Gisteren liep de overwinning van Ajax uit... op een totale chaos in de Amsterdamse binnenstad.",
  "transcript": "Gisteren liep de overwinning van Ajax uit.. op een totale chaos in de Amsterdamse binnenstad.",
  "transcript": "al",
  "transcript": "Als",
  "transcript": "van",
  "transcript": "voor",
  "transcript": "G",
  "transcript": "en",
  "transcript": "een",
  "transcript": "van clubs zijn. We gaan",
  "transcript": "van club zijn. We gaan",
  "transcript": "die van clubs zijn. We gaan",
  "transcript": "je staat",
  "transcript": "je staat.",
  "transcript": "en je staat",
  "transcript": "man in het zwart, petje op zak,",
  "transcript": "man in het zwart, petje op zak.",
  "transcript": "man in het zwart, petje op zak, R",
  "transcript": "spot is op pad en ik wacht op contact.",
  "transcript": "De spot is op pad en ik wacht op contact.",
  "transcript": "tas.",
  "transcript": "pas.",
  "transcript": "was.",
  "transcript": "geen staande.",
  "transcript": "geen staande",
  "transcript": "geen staande op.",
  "transcript": "concentraat.",
  "transcript": "concentraat",
  "transcript": "concentr.",
  "transcript": "we lopen.",
  "transcript": "lopen.",
  "transcript": "de",
  "transcript": "At",
  "transcript": "er",
  "transcript": "zit.",
  "transcript": "zit?",
  "transcript": "zit",
  "transcript": "kan een shirt heel snel uitdoen of wegdoen. trap echt je koptrap... en anders ga ik jullie dat hele kroegje aan kan.",
  "transcript": "kan een shirt heel snel uitdoen of wegdoen. trap echt je koptrap... en anders ga ik jullie dat hele kroegje hier aan de kan.",

```


Kapitel 2 - Das war nicht ganz überzeugend

Trotz der aufwändigen Vorbereitung des Materials konnten die Lyrics des Songs nur teilweise erkannt werden.



Auch mit der Hilfe von GPT4 in der Nachbearbeitung der Texte haben wir weiterhin nur eine sehr unvollständige Version - zusätzlich haben wir versucht, selbst noch ein paar Stellen rauszuhören und mit Google Translate zu übersetzen.

Step 1: Creating a Refined Version

Given the fragmented nature of the text and its repetition, we aim to create a smooth, logical progression for the rap, keeping the context and flow in mind. Here's a consolidated version:

```
swift
Gisteren liep de overwinning van Ajax uit op een totale chaos in de Amste
Van clubs zijn we gaan, je staat man in het zwart, petje op zak, de spot
Geen staande, concentreer, we lopen, er zit kan een shirt heel snel uitdo
En anders ga ik hier het hele kroegje aan de kan, onze vak van clubs, ge
Stille stap, mensen vechten nog binnen. Zoveel actie, het lijkt net een f
Jij bent met sukkel en ik ben met killers, richting de vakten er rell
Ik van Xanders heb gezien en niet hierzo. Vooral die mensen had ik echt m
Scanner, I saw that there is some attack.
```

This version aims to maintain the essence and energy of the original text while making it more coherent.

Step 2: Translation into English

Now, let's translate this refined version into English:

```
vlone
Yesterday, Ajax's victory turned into total chaos in the streets of Amste
From clubs, we go, you stand man in black, cap in pocket, the spotlight i
No standing, concentrate, we walk, it's possible to quickly take off or ti
Otherwise, I'm going to take this whole bar down, our area of clubs, the
Silent step, people are still fighting inside. So much action, it's just
You're with losers, and I'm with killers, heading towards the areas where
I've seen Xanders and not here. Especially those people, I would have pul
Scanner, I saw that there is some attack.
```

Deutsch ↔ Niederländisch

wir gehen wieder auf jagd × We gaan weer op jacht

In Google Übersetzer öffnen • Feedback geben

Kapitel 3 - Hm... GPT4...

Machen wir's doch wie üblich in Python. :D - Download auf unserem Github Account.

```
home > crunchy > bin > mspeechtrans > multi_speech_trans.py > atoi
1  #!/usr/bin/env python3
2
3  import os
4  import glob
5  import argparse
6  from pydub import AudioSegment
7  import speech_recognition as sr
8  from googletrans import Translator, LANGUAGES
9  import openai
10 import time
11 import glob
12 import re
13
14 # API Keys - ensure to fill these with your actual API keys where needed
15 OPENAI_API_KEY = None
16 GOOGLE_CLOUD_SPEECH_API_KEY = None
17 WIT_AI_API_KEY = None
18 AZURE_SPEECH_API_KEY = None
19 HOUNDIFY_CLIENT_ID = None
20 HOUNDIFY_CLIENT_KEY = None
21 IBM_SPEECH_TO_TEXT_API_KEY = None
22 IBM_SERVICE_URL = None
23
24 # Global Settings
25 RETRY_DELAY = 3
26 RETRY_COUNT = 3
27 translator = Translator(service_urls=['translate.googleapis.com'])
28 r = sr.Recognizer()
29 openai.api_key = OPENAI_API_KEY
30
31 def atoi(text):
32     return int(text) if text.isdigit() else text
33
34 def natural_keys(text):
35     return [ atoi(c) for c in re.split(r'(\d+)', text) ]
36
37 def get_audio_files(folder_path):
38     """Return a list of audio file paths from the given folder, sorted in human-readable ascending order."""
39     files_grabbed = []
40     for files in ('*.mp3', '*.wav'): # Extend or adjust as needed
41         files_grabbed.extend(glob.glob(os.path.join(folder_path, files)))
42
43     # Sort the files using the natural_keys function
44     files_grabbed.sort(key=natural_keys)
45
46     return files_grabbed
47
48 def transcribe_with_whisper(audio_path, rec_language='en'):
49     """Transcribe audio using OpenAI's Whisper model."""
50     try:
51         with open(audio_path, "rb") as audio_file:
52             transcript = openai.Audio.transcribe('whisper-1', audio_file)
53             if 'text' in transcript:
54                 return {"service": "Whisper", "text": transcript['text']}
55             else:
56                 return {"service": "Whisper", "error": "Transcription failed without specific error message."}
57     except Exception as e:
58         return {"service": "Whisper", "error": str(e)}
59
60 def transcribe_audio(file_path, rec_language):
61     """Transcribe audio using various services."""
62     # Convert to WAV if needed for compatibility
63     audio, wav_path = convert_to_wav(file_path)
64
65     results = {}
```


Kapitel 4 - Bessere Ergebnisse - und mehr

Nach einer langen Nacht ist das Script fertig und funktioniert. Wir haben bisher nur 3 der insgesamt 6 verschiedenen AI-Services getestet, die wir ins Script integriert haben, da man überall API Keys braucht. Unter anderem OpenAI Whisper hat uns deutlich mehr überzeugt.

```
$ mspeechtranscode /mnt/c/Users, /Desktop/audiotrans_project/ -t en -l nl

[***] Processing file: Test1.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test2.wav
[+] Transcribed using WHISPER

[***] Processing file: Test3.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test4.wav
[+] Transcribed using WHISPER

[***] Processing file: Test5.wav
[+] Transcribed using WHISPER

[***] Processing file: Test6.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test7.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test8.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test9.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test10.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test11.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test12.wav
[+] Transcribed using WIT API
[+] Transcribed using WHISPER

[***] Processing file: Test13.wav
[+] Transcribed using WHISPER

[***] Processing file: Test14.wav
```

Kapitel 4 - Bessere Ergebnisse - und mehr

Obendrauf haben wir noch Google Translate integriert. Das Script fasst die Ergebnisse zusammen und übersetzt. Alles automatisch, deutlich schneller, einfacher und billiger als bei AWS.

```
[+] Results from WHISPER:
Gisteren liep de overwinning van Ajax uit tot een totale chaos in de Amsterdamse binnenstad.
Stay Homie Agreeen ACL Stay Homie Da stay Homie Da stay Homie
I'll say it again. Say it. Go on. It's great. It's fun. Go around. Go to school. Have fun. Go everywhere. Go to
We gaan weer op jacht, honderd man in je staat Honderd man in het zwart, bitches op zak Panta's shit vacht, spo
Wacht op locatie, we gaan z'n taart zien. We klappen geen staande ovatie. Alleen maar chaos, escalatie. We zijn
Je kan de shirt heel snel uitdoen of wegdoen. En dan krijgt u een koptrap en dan gaat u er helemaal in genieren
Azzel A, Azzel A, Azzel A, 1-3-1-2, Fuck de Romeos en fuck de N.E. Rijlen in de stad of gevecht in café, we bre
The breakable limits for Crips and M.A.
Bollen gaan af, verringen van glas, storm naar binnen en handelen het af. Krukken de lucht tegen honden die bla
Mensen vechten nog binnen, zoveel actie het lijkt net een film Jij bent met sukkels en ik ben met killers
Peloton in m'n ooghoeken, moet nog effe instillen op z'n ooghoeken Laat hem in de rook naar z'n fucking oog zo
Vandaag was je aan, maar je ligt te gaan uit
ACA, 1-3-1-2, fuck de Romeo's en fuck de M.A. Rel in de stad of gevecht in café, we breken door linies van club
Als ik ergens anders had gezien en niet hier was gehoord, vooral die meisjes, dan had ik echt meteen mijn kop e

[+] Translation of WHISPER transcription to en:
Yesterday, Ajax's victory resulted in total chaos in the city center of Amsterdam.
Stay Homie Agreeen ACL Stay Homie Da stay Homie Da stay Homie
I'll say it again. Say it. Go on. It's great. It's fun. Go around. Go to school. Have fun. Go everywhere. Go to
We're going on the hunt again, a hundred men in your state, a hundred men in black, bitches in the pocket of Pa
Wait on location, we're going to see his cake. We don't give a standing ovation. Just chaos, escalation. We are
You can take off or throw away the shirt very quickly. And then you get a head kick and then you completely fre
Azzel A, Azzel A, Azzel A, 1-3-1-2, Fuck the Romeos and fuck the N.E. Driving in the city or fighting in a café
The breakable limits for Crips and M.A.
Bulbs go off, glass breaks, storm in and deal with it. Crutches the air against dogs that bark only quietly and
People are still fighting inside, so much action, it looks like a movie. You are with losers and I am with kill
Peloton in the corner of my eye, still need to take some instillation on the corner of his eye. Let him look fo
Today you were on, but you are about to go off
ACA, 1-3-1-2, fuck the Romeos and fuck the M.A. Riot in the city or fight in a café, we break through lines of
If I had seen anywhere else and not heard here, especially those girls, I would have blown my head off in a hea
```

Download: <https://github.com/Network-Sec/maudiotrans>