Api de reconnaissance émotionnelle

Février 2018

# Reconnaissance émotionnel Vocal

## BeyondVerbal V4 **(/!\ V5)**

“With over 21 years of research, close to two million voice samples, 18 scientific discoveries and a portfolio of U.S. - granted patents – it is our vision (and our mission) to endow machines with the ability to connect with us on an emotional level,”.

BEYONDVERBAL « au-delà des mots », société israélienne au-devant de la scène de la reconnaissance émotionnel dans la voix depuis 2007, propose un service de reconnaissance émotionnel dans des fichiers audios. Ce service basé sur une IA entrainé sur **2.3 Million de fichier audio de 170 pays, analysé par des algorithmes fils de 8 brevets et 21 ans de recherche**.

L’api prend en entrée **un fichier son d’au moins 10s, et après 5 à 20\* secondes, revoie un json** notant les caractéristiques suivantes :

* « Valence » : Mesure la positivité de la phrase
* « Arousual » : Estime si la voit est plutôt « tranquille / ennui » ou plutôt « vigilent / excité »
* « Temper » : Estime si la voit est plutôt « morose » < « amical » < « dominant / agressif »

L’api fourni également deux autres informations :

* ~~« Mood groups » : Une liste d’humeurs retrouvé dans la voix~~
* ~~« Combined emotions » : Une liste d’émotion retrouvé dans la voix. (Entre 7, 11 ou 21) ( les 11 :~~ **~~tristesse~~**~~, solitude, anxiété, neutre(self-control), amical, motivé, passionné, joyeux, dominant, critiquant, colère)\*\*~~
* « Emotion\_group » : le groupe d’émotion trouvé et le taux de confiance. (cf Annexes)

Cf : <http://www.beyondverbal.com/api/>

<http://www.beyondverbal.com/output-definition/>

Annexe : BeyondVerbal output

Ces deux derniers points sont le point fort de cette api, qui est de préciser les émotions du locuteur. En effet peu d’autre font ce travail supplémentaire qui est, en gros, d’analyser les trois premier afin d’en trouver une émotion. Ce qui facilite grandement le ciblage de l’émotion du locuteur.

La société permet de tester l’api pendant un mois puis une **version gratuite permet de faire des tests à raison de 100 minutes par mois** de son traité par mois.

Ci-dessous un exemple en « direct » de détection émotionnel :

<https://www.youtube.com/watch?v=jW9AZiZ3vmo>

Cette société semble proposer la meilleure qualité de service pour le moment.

\* : valeurs constatées.

\*\* : les groupe d’humeur vont être scoré dans une version attendu fin mars 2018 + group5

## Empath webApi :

« Empath identifies your emotion by analyzing not what you say but physical properties of your voice such as pitch, tone, speed, and power. We can identify your emotion whatever language you speak. »

Empath est le résultat des recherches en reconnaissance émotionnel de Smartmedical Corps : une entreprise japonaise de la santé.

Basé sur 10 milles fichier audio, Empath analyse les propriétés physiques de la voix afin d’en détecter cinq paramètres :

* « anger » : niveau de colère dans la voix
* « joy » : de joie
* « sadness » : de tristesse
* « calmness » : de calme
* « vigor » : de vigueur

Le fichier son en entrée doit être sous **un format bien précis** ( cf : annexes/EmpathDoc ) et d’une durée **inférieur à 5 secondes**.

L’api est disponible dans une version gratuite, étant payante, elle permet donc de la tester à hauteur de **300 requêtes par mois**.

Ci-dessous un exemple de ce que retourne l’api :

{"error":0, "anger":0, "joy":35, "sorrow":24, "calm":21, "energy":20}

Simple et efficace(?), l’api permet de faire ressortir les principales émotions mais on regrette de ne pas trouver de tableau de correspondance afin de pouvoir identifier, par exemple, le stress.

Vous trouverez en suivant le lien ci-dessous un certain nombre de projet comme « Jibun Yohou(My Mood Forecast) » qui est une application croisant donnée météo et humeur des employés afin d’y trouver une potentiel corrélation. Cette dernière s’appuie donc sur empath api, comme «NTT docomo Tohoku Reconstruction Project » ( projet médical ayant pour but de suivre l’état mental des victimes tremblement de terre) et d’autre projet. On retrouvera également le projet Utakata Mood-Light® en collaboration avec Philips Hue, un service d’ampoule RGB, le service va utiliser l’api empath afin de modifier l’ambiance lumineuse de la pièce en fonction de l’humeur ressentit.

Site web: <https://webempath.net/lp-eng/>

## Affectiva - speech

Projet sortie des labos du MIT en 2009, Affectiva se segmente en plusieurs corps de recherche dédié à la reconnaissance émotionnel : Visage, voix et texte.

Pour l'instant, seule la reconnaissance d’image : face, est opérationnel et commercialisé en tant que tel. Mais le reste arrive pour bientôt, étant actuellement accessible en échange de données.

L’api retournerais les infos suivantes sur le son envoyé :

* « Laughing » : l’action ou son de rire.
* « Anger/Iritation » : Forte expression de mal-être, hostilité, irritation ou frustration.
* « Arousual » : niveau de vigilance, excitement ou engagement envers le sujet de l’émotion.
* « gender » : détermine si la voix est plutôt féminine ou masculine.

Des caractéristiques peu communes aux autre api qui peuvent rapidement apporter un plus sur un site de vente, notamment le genre de la voix, qui sort cependant un peu du cadre « émotionnel ».

Une société qui vise cependant les gros projets à la vue d’une facturation débutant à 5000$.

## OpenSmile :

OpenSmile, par audEERING, est une librairie open source C++ qui permet d’extraire du son envoyé un grand nombre d’information : <https://audeering.com/technology/opensmile/#features>

Pour celle qui relié au texte :

* Signal energy
* Loudness
* Mel-/Bark-/Octave-spectra
* MFCC
* PLP-CC
* Pitch
* Voice quality (Jitter, Shimmer)
* Formants
* LPC
* Line Spectral Pairs (LSP)
* Spectral Shape descriptors

OpenSmile est difficile à interpréter mais semble poser de bonnes bases, utilisé dans différentes applications de audEERING, qui propose une application de reconnaissance émotionnel : sensAI.

Malheureusement, à cause des milliers d’homonyme, les recherches à ce sujet sont très difficiles et je n’ai pas trouvé d’information complémentaire sur les émotions reconnue (la présentation annone : joie, colère, peur, etc.) et/ou facturation. (<https://audeering.com/technology/sensai/>)

L’intérêt d’openSmile d’être open source et de pouvoir être ainsi intégré dans un service maison.

## GoodVibration :

* sdk ?
* <http://www.good-vibrations.nl/>

Vokaturi :

« The Vokaturi software reflects the state of the art in emotion recognition from the human voice. Its algorithms have been designed, and are continually improved, by Paul Boersma, professor of Phonetic Sciences at the University of Amsterdam, who is the main author of the world’s leading speech analysis software Praat. »

Vokaturi est une librairie permettant de détecter des émotions dans la voix. Disponible sous licence GPL ou plus, en C ou python, elle permet de détecter des émotions dans le spectre de la voix.

Elle s’appuie sur les quatre émotions suivantes (et l’absence ces dernières) :

* Joie
* Tristesse
* Colère
* Peur
* Neutre.

La version open-source annonce un taux de fiabilité de 66.5%, mais la libraire propose de monter jusqu’à 76.1%.

# Annexes :

## BeyondVerbal :

### output exemple : (V4 : attention à la màj)

{"status":"success","result":{"**duration**":"21304.25","sessionStatus":"Done","analysisSegments":[{"offset":0,"duration":10000,"end":10000,"analysis":{"**Temper**":{"Value":"93.00","Group":"high","Score":"90.00"},"**Valence**":{"Value":"16.00","Group":"negative","Score":"76.00"},"**Arousal**":{"Value":"86.00","Group":"high","Score":"80.00"},"**Vad**":{"Voiced":"75.00"},"**Mood**":{"**Group7**":{"Primary":{"Id":1,"Phrase":"Angry"},"Secondary":{"Id":1,"Phrase":"Angry"}},"**Group11**":{"Primary":{"Id":2,"Phrase":"Criticism,Cynicism"},"Secondary":{"Id":5,"Phrase":"Hostility, Anger"}}, "**Group21**": {"Primary":{"Id":7,"Phrase":"disliking"},"Secondary":{"Id":14,"Phrase":"hostility"}},"Composite":{"Primary":{"Id":153,"Phrase":"Sarcasm. Possibly restrained anger."},"Secondary":{"Id":111,"Phrase":"Loss of inhibitions, aggressiveness."}}}}}, {"offset":5000,"duration":10000,"end":15000,"analysis":{"**Temper**":{"Value":"96.00","Group":"high","Score":"94.00"},"**Valence**":{"Value":"29.00","Group":"negative","Score":"55.00"},"**Arousal**":{"Value":"88.00","Group":"high","Score":"82.00"},"**Vad**":{"Voiced":"72.00"},"**Mood**":{"**Group7**":{"Primary":{"Id":1,"Phrase":"Angry"},"Secondary":{"Id":1,"Phrase":"Angry"}},"**Group11**":{"Primary":{"Id":5,"Phrase":"Hostility, Anger"},"Secondary":{"Id":11,"Phrase":"Supremacy, Arrogance"}},"**Group21**":{"Primary":{"Id":14,"Phrase":"hostility"},"Secondary":{"Id":14,"Phrase":"hostility"}},"Composite":{"Primary":{"Id":75,"Phrase":"Aggressive urges."},"Secondary":{"Id":117,"Phrase":"Openly confrontational."}}}}}, {"offset":10000,"duration":10000,"end":20000,"analysis":{"**Temper**":{"Value":"99.00","Group":"high","Score":"98.00"},"**Valence**":{"Value":"23.00","Group":"negative","Score":"65.00"},"**Arousal**":{"Value":"99.00","Group":"high","Score":"98.00"},"**Vad**":{"Voiced":"69.00"},"**Mood**":{"**Group7**":{"Primary":{"Id":1,"Phrase":"Angry"},"Secondary":{"Id":1,"Phrase":"Angry"}},"**Group11**":{"Primary":{"Id":11,"Phrase":"Supremacy, Arrogance"}, "Secondary":{"Id":11,"Phrase":"Supremacy, Arrogance"}},"**Group21**":{"Primary":{"Id":8,"Phrase":"dominance"}, "Secondary":{"Id":8,"Phrase":"dominance"}}, "Composite":{"Primary":{"Id":36,"Phrase":"Confrontational. Attempt to control/dominate."},"Secondary":{"Id":399,"Phrase":"Fanatical preaching. Authoritative."}}}}}], "analysisSummary":{"AnalysisResult":{"**Temper**":{"Mode":"high","ModePct":"100.00"},"**Valence**":{"Mode":"negative","ModePct":"100.00"},"**Arousal**":{"Mode":"high","ModePct":"100.00"}}}},"recordingId":"8ef5dc3d-95bf-4ec3-a6f8-96ee9d4ceee7"}

### Etude de la réponse :

On retrouve dans chaque réponse les trois principaux paramètres : temper, valence, arousual. Mais surtout ce qui va être intéressant : les émotions/humeurs.

Le temps de son analyser étant de 10 secondes, afin de préciser l’humeur sur la durée de la conversation, l’api découpe le son en tranche de 10s avec 5s entre chaque début.

Sur l’exemple si dessus, le son de 21s est ainsi découpé :

* 0 -> 10
* 5 -> 15
* 10->20

Chaque sous partie aura son analyse et on pourra en sortir tous les détails de l’appli.

### Groupe d’émotions : (V4)

3.1 AGGRESSIVE / CONFRONTATIONAL MOOD GROUPS

* Supremacy and Arrogance.
* Hostility and Anger.
* Criticism and Cynicism.

3.2 SELF-CONTROL MOOD GROUP

* Self-control and practicality.

3.3 EMBRACIVE MOOD GROUPS

* Leadership and Charisma.
* Creativeness and Passion.
* Friendliness and Warm.
* Love and Happiness.

3.4 DEPRESSIVE / GLOOMY MOOD GROUPS

* Loneliness and Unfulfillment.
* Sadness and Sorrow.
* Defensiveness and Anxiety.

### Groupe d’émotion : (V5)

* Sadness/Uncertainty/Boredom
* Anger/Dislike/Stress
* Neutral
* Happiness/Enthusiasm/Friendliness
* Warmth/Calmness

## EmpathDoc :

