

API V4 VOICE INPUT GUIDELINES

Voice Input Guidelines for Analysis by Beyond Verbal's REST API

VOICE SAMPLE COLLECTION

Beyond Verbal's REST API (API) supports both near real time voice analytics on ongoing uploads of consecutive chunks of the recorded signal, as well as offline voice analytics of prerecorded voice samples.

Our Emotions Analytics engine requires a **minimum 13 seconds of recorded speech** before first analysis response. After these initial 13 seconds, a new analysis response is generated every 4 seconds.

Note, the analysis is performed simultaneously during the upload process, ensuring a rapid response once enough voice data has been collected

Required Input format

- The API requires a codec of WAV PCM 8 KHz, 16-bit Mono.
- In theory, the API can accept voice signals from either audio or video that were converted to the above required format via suitable codecs / resampling techniques.
- In the event that you need to convert voice signals we recommend using ffmpeg, it is a free open source available at: <https://ffmpeg.org/ffmpeg.html> an example on how to convert to our supported format is: `ffmpeg -i input_file.wav -acodec pcm_s16le -ac 1 -ar 8000 output_file.wav`

Signal quality

- Please be aware, that the signal quality affects the performance of the emotion recognition. Thus, it is NOT recommended to use low quality recording devices nor to use signals that were decoded from high compression encoders such as those used for Voice over IP.
- Avoid (as much as possible) using signals that are saturated and/or clipped as the recognition engine examines very fine elements in the signal that get lost in such cases.



Saturated voice



Normal voice

- Poorly recorded voice inputs contain voice deformities, excessive hiss and other vocal contaminants. Such recorded samples are inappropriate for analysis. *For example, Inputs from mobile phones generally tend to produce higher-quality voice samples than laptops.*

- The voice input should only consist of a single speaker at a time. Having multiple speakers on the same voice input will skew the analysis.
- Although our algorithms are – to an extent – noise tolerant, the lower the background noise, the higher the accuracy of the analysis. Try recording in a moderately quiet room, with relatively low background noise (TV, other people speaking, fans, noisy external city sounds etc.).

Not receiving an analysis?

Here are some common problems and easy fixes?

- Not enough voice to analyze.
- Recording in a car or similarly noisy environment (such as planes, restaurants, trains, and automobiles). Try again at home or office.
- Sitting too close or out of reach of the microphone. A good distance is to position yourself about 15-50 cm (6 – 20 inches) from the microphone to avoid sound saturation and reduce white noise.
- Speaking too loud or too soft. Speak as you would in a normal conversation—clearly, but no need to be extremely loud.
- More than one person is heard in the recording (this can also result from a noisy TV in the background). Wait for a quiet moment, and make sure your TV and radio are off.
- Using halting speech. Imagine you are speaking to a friend.