call journey

AI Powered Voice Analytics

wordbench

# Json Output Format Guide

May 2020

call journey
AI Powered Voice Analytics

# Contents

# 1. OFFICIAL DISCLAIMER

THE INFORMATION CONTAINED IN THIS DOCUMENT IS THE PROPRIETARY AND CONFIDENTIAL INFORMATION OF VOICEAI PTY LTD (VOICEAI). YOU MAY NOT DISCLOSE, PROVIDE OR MAKE AVAILABLE THIS DOCUMENT, OR ANY INFORMATION CONTAINED IN THIS DOCUMENT, TO ANY THIRD PARTY, WITHOUT THE PRIOR WRITTEN CONSENT OF VOICEAI.

The Information in this document is provided for use with VoiceAI products. No license, express or implied, to any intellectual property associated with this document or such products is granted by this document.

All VoiceAI products described in this document are owned and/or licensed by VoiceAI and are protected by patents, trade secrets, copyrights or other intellectual/industrial property rights.

The VoiceAI products described in this document may still be in development. The final form of each product and release date thereof is at the sole and absolute discretion of VoiceAI. Your purchase, license and/or use of VoiceAI products shall be subject to VoiceAI' then current terms and conditions.

Trademarks

The following are trademarks of VoiceAI in Australia and other countries:

- Call Journey
- VoiceAI
- WordBench
- Record Retrieve

# 2. OVERVIEW

## Intended Audience

This manual is written for developers looking to parse the output JSON from the WordBench product.

## About the JSON file format

VoiceAI' speech-to-text engines handle speech to text conversion and provide additional metadata of each processed audio file including emotional intelligence, confidence scores, word start and end times, and other call information. VoiceAI products use the JSON file format to store the text information in an organized fashion to provide the most content that can be derived from the audio source. Products such as WordBench may have additional JSON elements that are not included in the base JSON structure.

JSON stands for JavaScript Object Notation, which is a syntax for storing and exchanging data. It provides a human-readable collection of data that can be accessed in a concise and logical manner. JSON filename extension is **.json**.

## Additional Resources

The examples contained in this document are screenshots of the JSON Viewer software, which can be downloaded for free at the following link: https://jsonviewer.codeplex.com/

# 3. JSON OUTPUT SCHEMA

Before defining each JSON element, it should be noted that when a data value is not defined (null), the data name will not appear as an element in JSON output, so keep in mind that, when viewing a specific JSON output file, not all the elements listed below may be included.

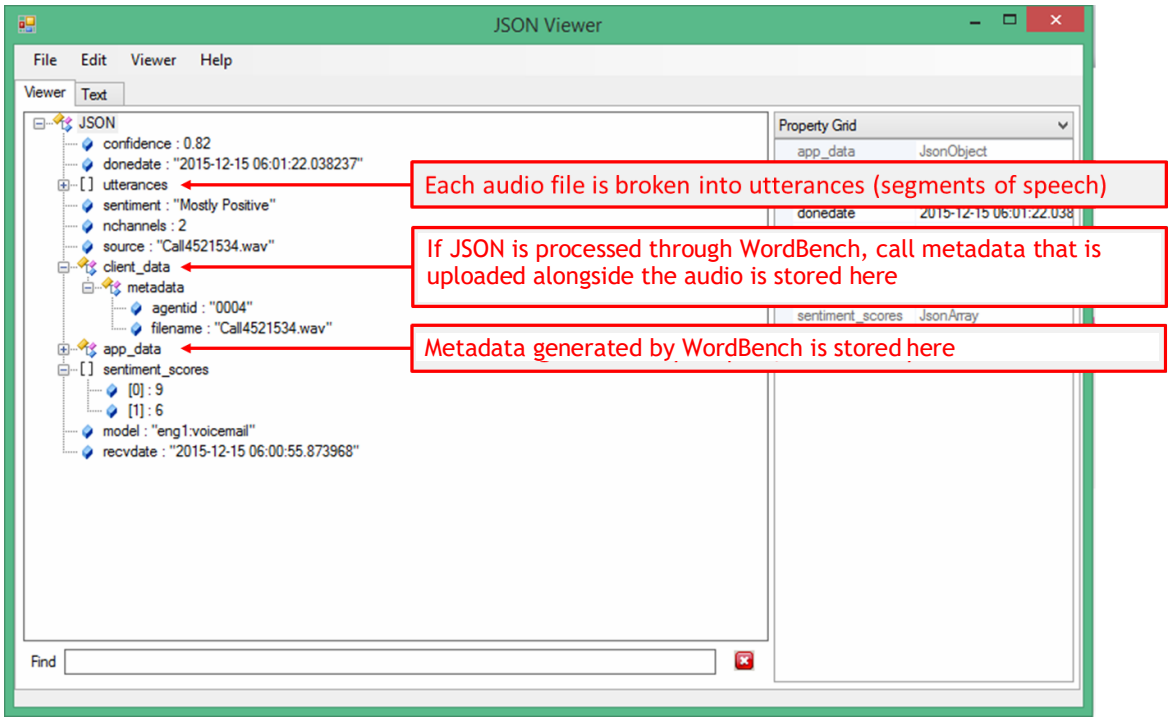
## Top-level Elements

The top-level elements included in a VoiceAI JSON are described below.

| Element | Type | Definition |
|---|---|---|
| *app_data | object | JSON object that stores metadata generated by WordBench |
| *client_data | object | JSON object that stores call metadata associated with the audio file |
| confidence | value | A measure of how confident the speech recognition system is in its transcription results<br>• Range between 0 and 1<br>• 1 is most confident |
| donedate | value | Date and time the file transcription was completed by the speech-to-text engine |
| model | array | Language model(s) used to process this audio file |
| nchannels | value | Number of channels in the audio file unless diarization is set to true, in which a single (1) channel file is broken up into 2 based on speaker separation |
| recvdate | value | Date and time the audio file was received by the speech-to-text engine |
| *scrubbed | value | If true then audio is purified so numbers are all redacted. If false, the data name does not appear in the JSON output. |
| *sentiment | value | Linguistic sentiment value:<br>• Positive<br>• Mostly Positive<br>• Neutral<br>• Mostly Negative<br>• Negative<br>• Mixed (contains both Positive and Negative in the file) |
| *sentiment_scores | array | Array of length 2. [0]=Positive phrase counts and [1]=Negative phrase counts in the file |
| source | value | Audio file name |
| utterances | array | Contains the word transcripts and corresponding metadata organized by utterances (speech segments broken up by silence boundaries) |

*Indicates JSON elements that are only available with WordBench.

An example of the top-level information is shown below.




## The “utterances” Array

The top-level **utterances** element contains an array of speech segment information. Each array element consists of the following elements:

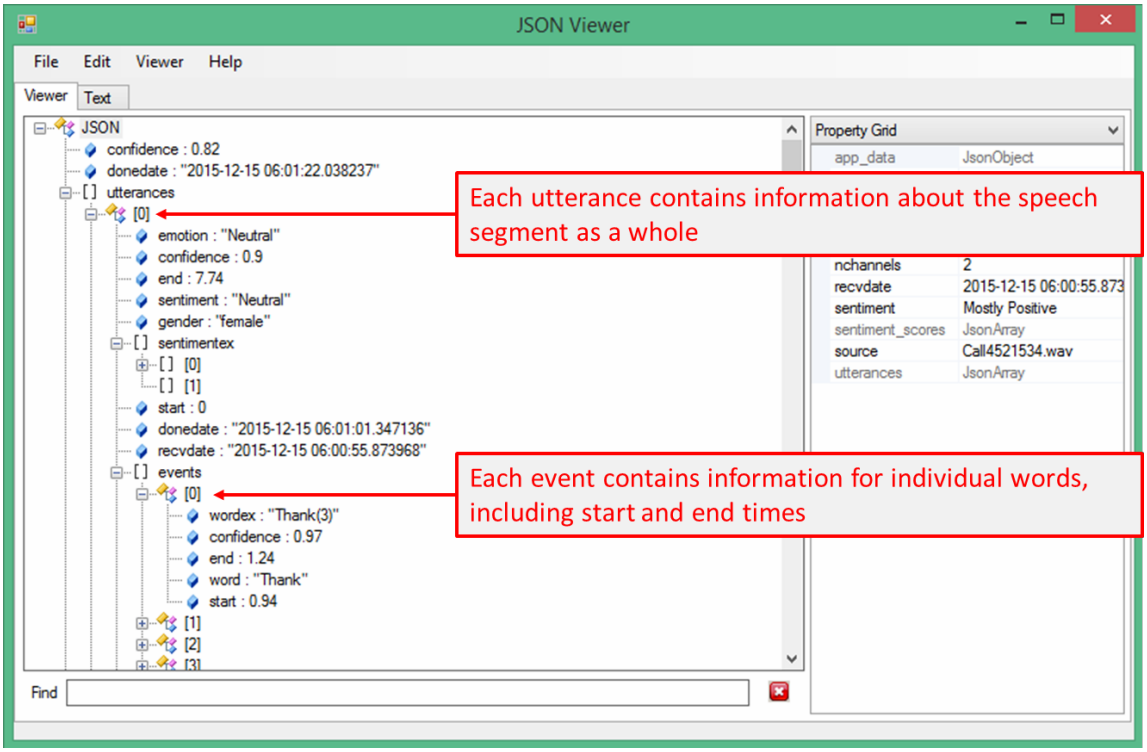
| Object | Type | Definition |
|---|---|---|
| metadata | object | Speaker information of the utterance. Each object contains the following values:<br>• **channel:** channel number<br>• **model:** model that decoded the utterance<br>• **source:** audio file name<br>• **uttid:** utterance segment number |
| confidence | value | A measure of how confident the speech recognition system is in its utterance transcription results<br>• Range between 0 and 1<br>• 1 is most confident |
| donedate | value | Date and time the utterance transcription was completed by the speech-to-text engine |
| *emotion | value | Emotional intelligence consists of both acoustic and linguistic information. Events can be given the following values:<br>• Positive<br>• Mostly Positive<br>• Neutral<br>• Mostly Negative<br>• Negative |
| end | value | End time of the utterance in seconds |
| events | array | Contains information about individual words. Each element is a word object that contains the following values:<br>• **confidence:** word level transcription confidence value between 0 and 1<br>• **end:** end time of the word in seconds<br>• **start:** start time in seconds<br>• **word:** normalized word<br>• **wordex:** raw dictionary word |
| *gender | value | Gender prediction of the speaker |
| recvdate | value | Date and time the utterance was received by the speech-to-text engine |
| *sentiment | value | Utterance-level linguistic sentiment value:<br>• Positive<br>• Mostly Positive<br>• Neutral<br>• Mostly Negative<br>• Negative<br>• Mixed (contains both Positive and Negative in the file) |

| Object | Type | Definition |
|---|---|---|
| *sentimentex | array | Contains sentiment information for each utterance<br>• [0][0]=Positive phrase counts<br>• [0][1]=Negative phrase counts in utterance<br>• [1][*] consist of an array of sentiment segments where [0] = '+' or '-' for Positive and Negative, and [1] is the position range of the phrase |
| start | value | Start time of the utterance in seconds |

*Indicates JSON elements that are only available with WordBench.

Each **utterances** element contains an **events** array, which provides information about each word in the utterance.

*Examples of an utterances array element and an events array are shown below:*



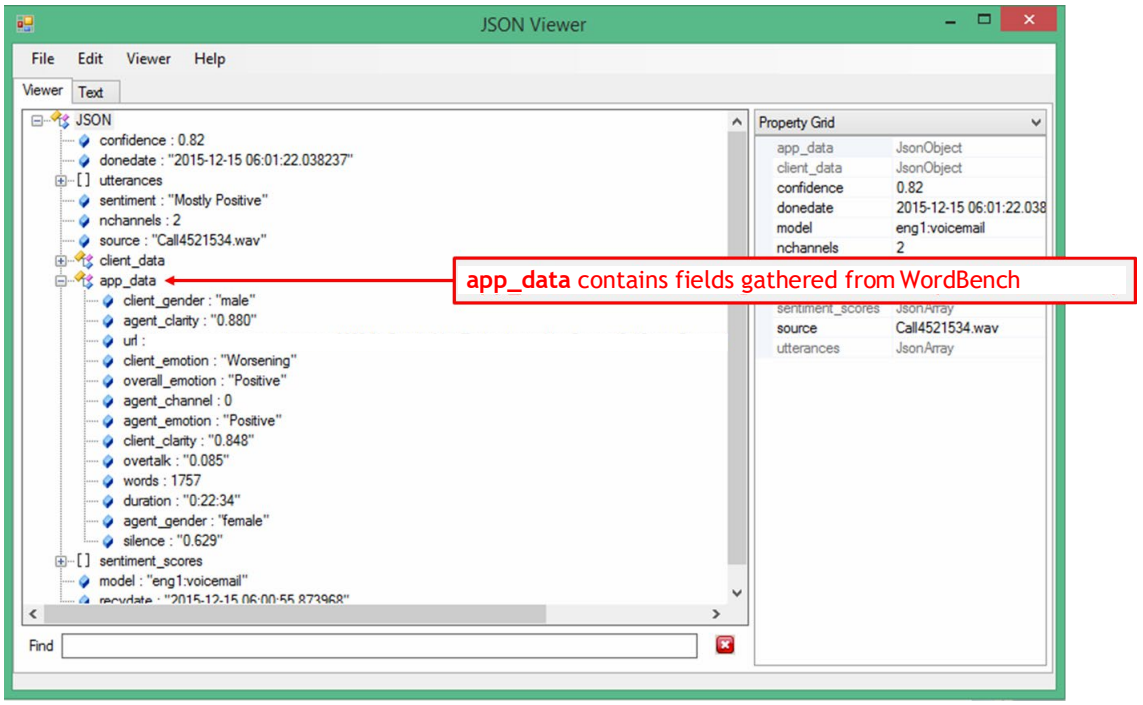

## The “app_data” Object

Another top-level JSON object is **app_data**. This object is **only** available when extracting JSON from WordBench. Metadata in this object applies to the whole call, rather than to individual utterances.

| Data Value | Definition |
|---|---|
| agent_channel | The channel the agent is on |
| agent_clarity | How clear the agent channel/speech is<br>• Range between 0 and 1<br>• 1 is clearest. |
| agent_emotion | Overall agent emotional intelligence consists of both acoustic and linguistic information and can have the following values:<br>• Positive<br>• Improving<br>• Negative<br>• Worsening |
| agent_gender | Agent gender prediction |
| client_clarity | How clear the client channel/speech is<br>• Range between 0 and 1<br>• 1 is clearest |
| client_emotion | Overall client emotional intelligence consists of both acoustic and linguistic information and can have the following values:<br>• Positive<br>• Improving<br>• Negative<br>• Worsening |
| client_gender | Client gender prediction |
| diarization | This value provided in 2 speaker, 1 channel calls.<br>• Range between 0 and 1.<br>• 1 is best speaker separation. |
| duration | Call duration |
| overall_emotion | Overall file emotional intelligence consists of both acoustic and linguistic information and can have the following values:<br>• Positive<br>• Improving<br>• Negative<br>• Worsening |
| overtalk | Percentage of call when the agent talks over or interrupts the client. Equal to the number of turns where the agent initiated overtalk divided by the total number of agent turns. |
| scorecard | Contains any application scores that have been calculated for the transcript |
| silence | Percentage of overall duration that is silence. Equal to all non-speech time, calculated as call duration minus the sum of the duration of each word. If music and noise is not decoded to word-events, they would be counted as silence. |
| url | Location in WordBench where the files can be viewed |

An example of **app_data** is shown below.

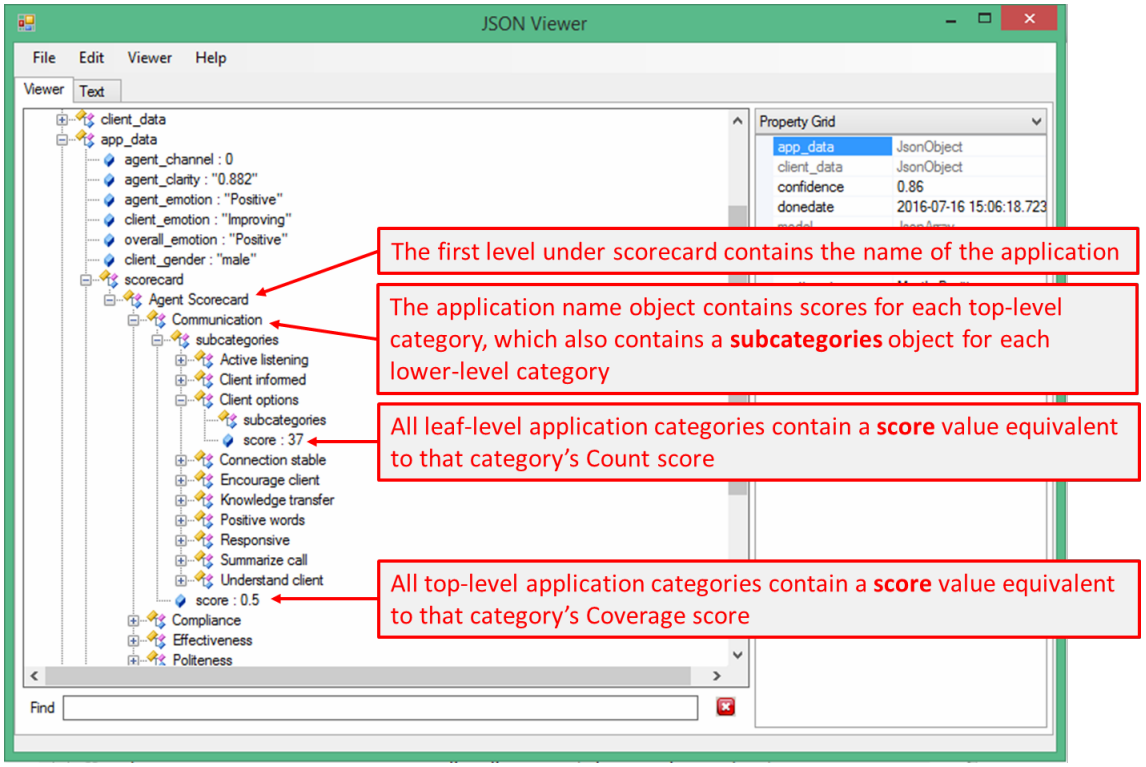


## The "scorecard" Object

Some JSON files processed in WordBench include a **scorecard** object under **app_data**. This object contains the Count and Coverage scores for each category in each application. For more information about how these scores are calculated and when they are included in a JSON output file, please refer to the WordBench Application Development Guide.

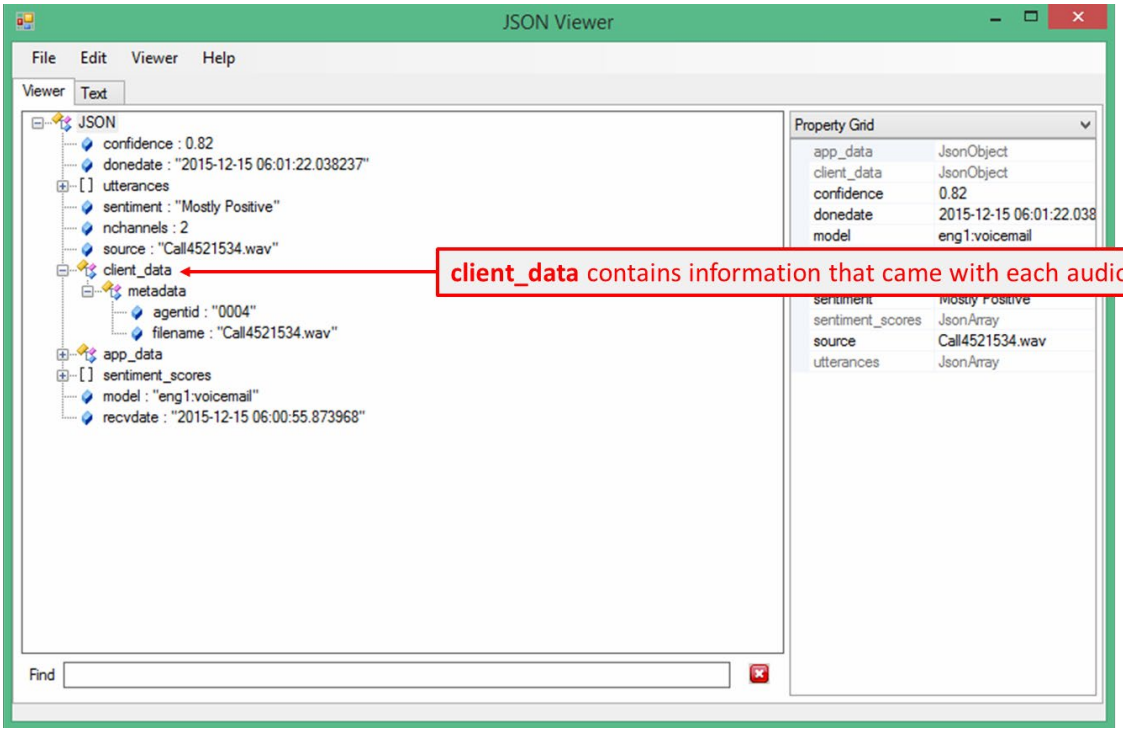
An example of **scorecard** is shown below.




## The "client_data" Object

Another top-level object that is **only** available with WordBench is **client_data**, which consists of one **metadata** object. This data is just a copy of the JSON or XML input metadata file uploaded to WordBench alongside the audio file. There is no fixed data name, and it is up to the user to define whether they want to pass any information into the output JSON files. An example output is shown below.



call journey

AI Powered Voice Analytics

VoiceAI

Powered By call journey