

# Talk Ratio Research Documentation

## Introduction

The objective of this research is to explore and compare readily available solutions for calculating the "Talk Ratio" at both an audio-based tonal level and an NLP-based level. Talk Ratio is defined as the percentage of talk time of Speaker 1 vs. Speaker 2 out of the total talk time. This research aims to identify open-source algorithms, and published, or commercialized third-party solutions with API integrations for this task. Furthermore, the research will provide a comparative analysis of these solutions, considering their pros and cons, projected accuracy, and costs. The projections are based on a data volume of 100,000 minutes.

## Audio-Based Tonal Level

### Solution 1: OpenSMILE

Description: OpenSMILE is an open-source audio feature extraction toolkit that can be used to extract various tonal and acoustic features from audio data.

Pros:

- Open-source and free to use.
- Extensive feature extraction capabilities.

Cons:

- Requires expertise in feature engineering and configuration.
- Limited out-of-the-box support for Talk Ratio calculations.
- Projected Accuracy: Moderate to high, depending on feature selection and configuration.
- Costs: Low (open-source).

### Solution 2: Third-Party API - IBM Watson Speech to Text

Description: IBM Watson offers a Speech to Text service that can transcribe audio data into text, which can then be processed to calculate Talk Ratio.

Pros:

- Well-documented and reliable.
- Easy integration with API.

Cons:

- Costs associated with API usage.
- Need for additional processing to calculate Talk Ratio.

Projected Accuracy: High for speech-to-text transcription, but depends on subsequent NLP analysis.

Costs: Moderate (based on API usage).

## NLP-Based Level

## Solution 1: spaCy

Description: spaCy is an open-source NLP library that can be used for natural language processing tasks, including text analysis and entity recognition.

Pros:

- Robust and highly customizable.
- Open-source and well-documented.

Cons:

- May require custom model training for Talk Ratio analysis.
- Learning curve for non-NLP experts.

Projected Accuracy: High, but model training and customization are needed.

Costs: Low (open-source).

## Solution 2: Third-Party API - Google Cloud Natural Language API

Description: Google Cloud's Natural Language API offers sentiment analysis and entity recognition, which can be leveraged for Talk Ratio analysis from text.

Pros:

- Easy API integration.
- High-level services for NLP.

Cons:

- Costs associated with API usage.
- Limited to available services (sentiment analysis and entity recognition).

Projected Accuracy: High for sentiment analysis, but Talk Ratio analysis may require additional processing.

Costs: Moderate (based on API usage).

## Comparative Analysis

Pros

- Audio-Based Solutions:

- OpenSMILE and IBM Watson Speech to Text are cost-effective (open-source or moderate API costs).
- IBM Watson Speech to Text provides accurate speech-to-text transcription.

- NLP-Based Solutions:

- spaCy and Google Cloud Natural Language API offer robust NLP capabilities.
- spaCy is open-source and highly customizable.

Cons

- Audio-Based Solutions:

- OpenSMILE requires expertise in feature engineering.
- IBM Watson Speech to Text necessitates additional processing for Talk Ratio calculation.

- NLP-Based Solutions:
  - Custom model training and learning curve (spaCy).
  - Limited services and API costs (Google Cloud Natural Language API).

#### Projected Accuracy

- Audio-Based Solutions have high accuracy in speech-to-text transcription but may vary for Talk Ratio.
- NLP-Based Solutions are highly accurate for sentiment analysis but may require additional processing for Talk Ratio.

#### Costs

- Audio-Based Solutions are cost-effective with low to moderate costs.
- NLP-Based Solutions involve moderate costs due to API usage.

#### Conclusion

The choice between audio-based and NLP-based solutions for Talk Ratio analysis depends on specific requirements and expertise. While audio-based solutions may provide direct speech-to-text transcription, NLP-based solutions offer robust NLP capabilities. Each solution has its pros and cons, and the decision should consider factors like cost, accuracy, and the complexity of implementation.

For a detailed implementation, testing, and cost analysis, it is advisable to prototype and evaluate these solutions with a representative dataset to determine the most suitable approach for the task.

This research documentation provides a comprehensive comparison of available solutions for Talk Ratio analysis at both audio-based tonal and NLP-based levels. It helps stakeholders make informed decisions based on their specific project requirements and constraints.