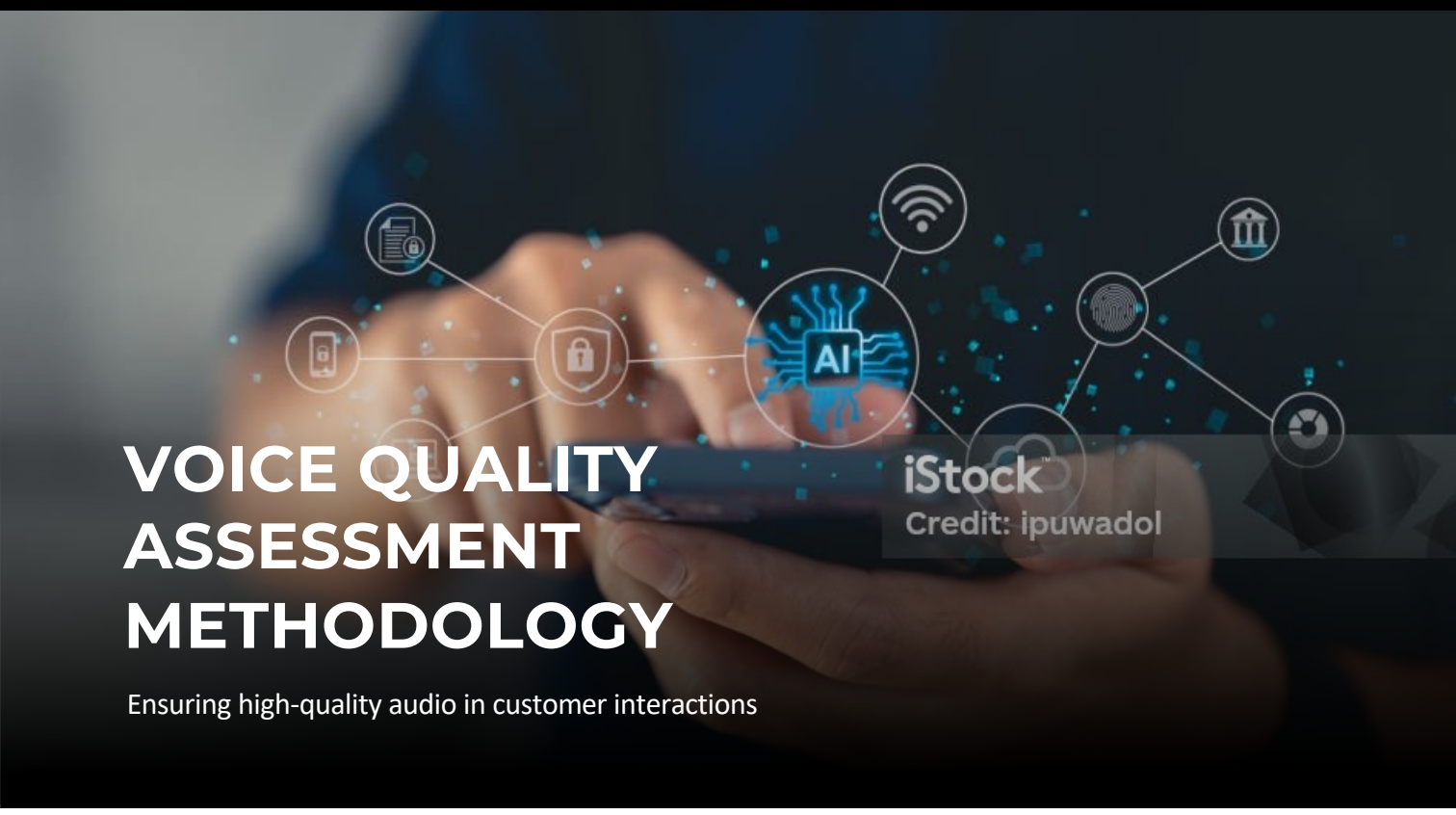

VOICE QUALITY ASSESSMENT METHODOLOGY

Ensuring high-quality audio in customer interactions

How to best assess voice quality of customer calls?

There are various methods employed to estimate the quality of voice and ensure that customer interactions are smooth. Mean Opinion Score (MOS) has been around for a while, but can Machine Learning help deliver a more dependable measure of voice clarity and intelligibility?

We think so.

MOS as measured by CCaaS providers

The Mean Opinion Score (MOS) provided by CCaaS providers is a measurement averaged over a large number of calls. It serves as an overall indicator of audio quality across multiple interactions, but due to its broad scope, it is not suitable for assessing the quality of individual calls. MOS in CCaaS providers focuses on long-term voice quality trends, making it useful for large-scale assessments, but less effective for single-call evaluations.

Transcript similarity based quality estimation

This method estimates voice quality by comparing the transcript of the actual call to a reference transcript. The higher the similarity between the two, the better the voice quality. This approach works by aligning the spoken words with the expected conversation, accounting for factors like timing and minor interruptions. The similarity score ranges from 0% to 100%, with higher percentages indicating better quality. This score is then converted into a scale of 1 to 5 for easier interpretation.

The Flowstate Machine Learning based MOS estimation

Our method has been developed to use advanced neural networks to estimate the MOS directly from audio recordings. Two specific neural nets are used in this approach:

- A neural network trained to filter out audio that contains no human speech. This ensures that only relevant parts of the recording are evaluated for voice quality.
- A neural network trained to assess speech comprehension across multiple languages. This allows for accurate assessment even in multilingual environments, ensuring that comprehension and clarity are evaluated.

These neural networks are part of the Flowstate speech-to-text transcriber and voice biometrics engine providing a passive, but highly accurate, measure of voice quality. They have been trained on over 680,000 hours of audio, covering more than 20 languages, making them a reliable choice for MOS estimation.

The final MOS score is computed by applying logarithmic transformations to mimic the way the human brain perceives sound quality. This score is then normalized to fall within a range of 0% to 100% and converted to a scale of 1 to 5.

© Omni Intelligence Pty. Ltd. 2025