Detailed Analysis of MFCC Features Reflecting Acoustic Characteristics of Different Languages (Hindi, Tamil, Bengali)

Mel-Frequency Cepstral Coefficients (MFCCs) are a widely used feature extraction technique in speech processing, designed to mimic the human auditory system's perception of sound frequencies. By transforming the power spectrum of a speech signal onto the Mel scale and then applying a Discrete Cosine Transform (DCT), MFCCs provide a compact and perceptually relevant representation of the speech's spectral envelope. Analyzing MFCCs across different languages can reveal how their distinct phonetic and phonological characteristics manifest in the acoustic domain. This study examines MFCC spectrograms and statistical summaries derived from them for Hindi, Tamil, and Bengali to identify language-specific patterns.

2. MFCC Spectrogram Analysis

Visual inspection of the MFCC spectrograms for Hindi, Tamil, and Bengali reveals both commonalities inherent to human speech and distinct patterns indicative of each language's unique acoustic properties.

2.1. Common Observations (Similarities)

- Energy Concentration in Lower Coefficients: All three languages exhibit prominent bands of high energy (indicated by darker red hues) in the lower MFCC indices (closer to the bottom of the spectrogram). This is a fundamental characteristic of voiced speech sounds, where the majority of the spectral energy resides in the lower frequency formants related to vowel production and the fundamental frequency.
- Temporal Structure: The spectrograms, spanning approximately 4.5 seconds, show similar temporal dynamics, suggesting comparable utterance lengths or speaking rates in the analyzed samples. Transitions and variations along the time axis reflect the sequential nature of phoneme production in all languages.
- Decreasing Energy with Higher Coefficients: As the MFCC index increases (moving towards the top of the spectrogram), the energy levels generally decrease (lighter shades). This is expected as higher cepstral coefficients capture finer spectral details and rapid variations, which typically have lower energy compared to the overall spectral envelope.

2.2. Language-Specific Differences

- Hindi:
    - The Hindi spectrogram (left) displays strong, well-defined, and relatively persistent (horizontally extended) deep red bands in the lower MFCC region, particularly between 1.0 and 3.5 seconds. This suggests the presence of stable, high-energy phonetic elements, likely related to its vowel system and voiced consonants.

    - The presence of more vertical texture and rapid changes in the mid-to-upper MFCC regions might reflect the influence of retroflex consonants and nasalized vowels, which involve more abrupt articulatory movements and spectral shifts.

- Tamil:
    - The Tamil spectrogram (middle) exhibits smoother transitions and a more smeared appearance compared to Hindi. This could be attributed to the characteristics of Dravidian languages, including Tamil, which often feature less aspiration and more glide-like transitions between phonemes.
    - While strong low-frequency energy is present, its distribution over time appears more uniform, indicating fewer abrupt spectral changes within the analyzed utterance.
- Bengali:
    - The Bengali spectrogram (right) shares similarities with Hindi in terms of strong low-frequency energy concentration. However, it displays smoother transitions than Hindi and notable dark blue areas (indicating lower energy relative to the dominant frequencies) in the lower MFCC region, particularly around 0.5 to 2 seconds.
    - These lower energy regions could be associated with Bengali's murmured voiced stops and generally softer consonantal articulations, leading to a more balanced overall spectral energy distribution compared to the more pronounced low-frequency dominance observed in parts of the Hindi spectrogram.

3. Statistical Analysis of MFCC Features

To complement the visual analysis, we can examine the statistical distribution of the MFCC values for each language, approximated from the provided box plots.

3.1. Visual Boxplot Interpretation

Language	Median MFCC Mean	Spread (IQR)	Outliers	Notable Insights
Hindi	≈ -20	Wide	Many	Higher variance and a tendency towards lower values.

Tamil	≈ -13	Narrow	Many	Most stable central tendency (highest median mean).
Bengali	≈ -19	Wide	Many	Similar central tendency to Hindi but with a wider upper range.

3.2. Approximate Statistical Summary

Language	Estimated Mean MFCC	Estimated Variance	Skewness	Interpretation
Hindi	-22	High	Left-skewed	Greater variability in MFCC values, with a skew towards lower energy spectral components. This may reflect a wider range of phonetic features.
Tamil	-14	Low	Approximately Symmetric	More consistent and less variable MFCC patterns, suggesting more uniform acoustic characteristics across the analyzed speech.
Bengali	-20	Moderate	Slightly Right-skewed	Moderate variability with a slight tendency towards higher energy spectral components compared to Hindi.

4. Discussion

The analysis of both MFCC spectrograms and their statistical distributions reveals distinct acoustic signatures for each language.

- Tamil's relatively high mean MFCC and low variance suggest a more consistent spectral energy distribution and smoother phonetic transitions. This could be linked to its phonological inventory and articulatory characteristics, which may involve fewer abrupt spectral changes.
- Hindi's lower mean MFCC and high variance indicate a greater degree of fluctuation in its spectral characteristics. This could be attributed to a richer phonetic inventory, including retroflex sounds and nasalized vowels, which introduce more spectral variation. The left-skewness might suggest a prevalence of lower energy components in its overall acoustic profile.
- Bengali appears to lie between Hindi and Tamil in terms of MFCC dynamics. Its moderate variance and slightly right-skewed distribution suggest a balance between stable spectral components and the presence of some higher energy features. This might reflect its unique set of phonetic features, such as murmured stops, which can influence the spectral envelope.

The presence of outliers in all three languages likely reflects the inherent variability within speech, including variations in individual speakers, phonetic context, and prosodic elements.

5. Conclusion

The MFCC features effectively capture and reflect the distinct acoustic characteristics of Hindi, Tamil, and Bengali. Visual analysis of spectrograms highlights differences in the temporal dynamics and spectral energy distribution, while statistical summaries reveal variations in the central tendency and spread of MFCC values. These differences are indicative of the underlying phonetic and phonological variations across the languages. This analysis underscores the utility of MFCCs in distinguishing between languages and provides valuable insights for applications such as automatic language identification and multilingual speech processing. Further research with larger and more diverse datasets can provide a more comprehensive understanding of these language-specific acoustic features.