Emotion Analysis Report: Valence-Arousal Analysis

Introduction

This report presents a comprehensive analysis of emotional content in the AMI Corpus dataset, specifically focusing on the ES2016a session. The analysis combines data from four video streams and one audio stream to provide a multi-modal perspective on emotional dynamics during the interaction.

Models and Parameters

Video Analysis Model

Model Type	Facial Expression Recognition
Framework	OpenCV + Custom Emotion Classifier
Features	Facial landmarks, expression patterns
Output	Valence (0-1), Arousal (0-1)
Frame Rate	30 fps
Processing	Frame-by-frame analysis with face detection

Audio Analysis Model

Model Type	Audio Feature Extraction + Emotion Recognition
Framework	Librosa + Custom V-A Predictor
Features	MFCC, Spectral Centroid, RMS Energy, Tempo
Output	Valence (0-1), Arousal (0-1)
Sampling Rate	3-second segments
Processing	Feature extraction followed by regression

Analysis Thresholds

The analysis uses the following thresholds to identify notable emotional moments:

Video Analysis Thresholds:

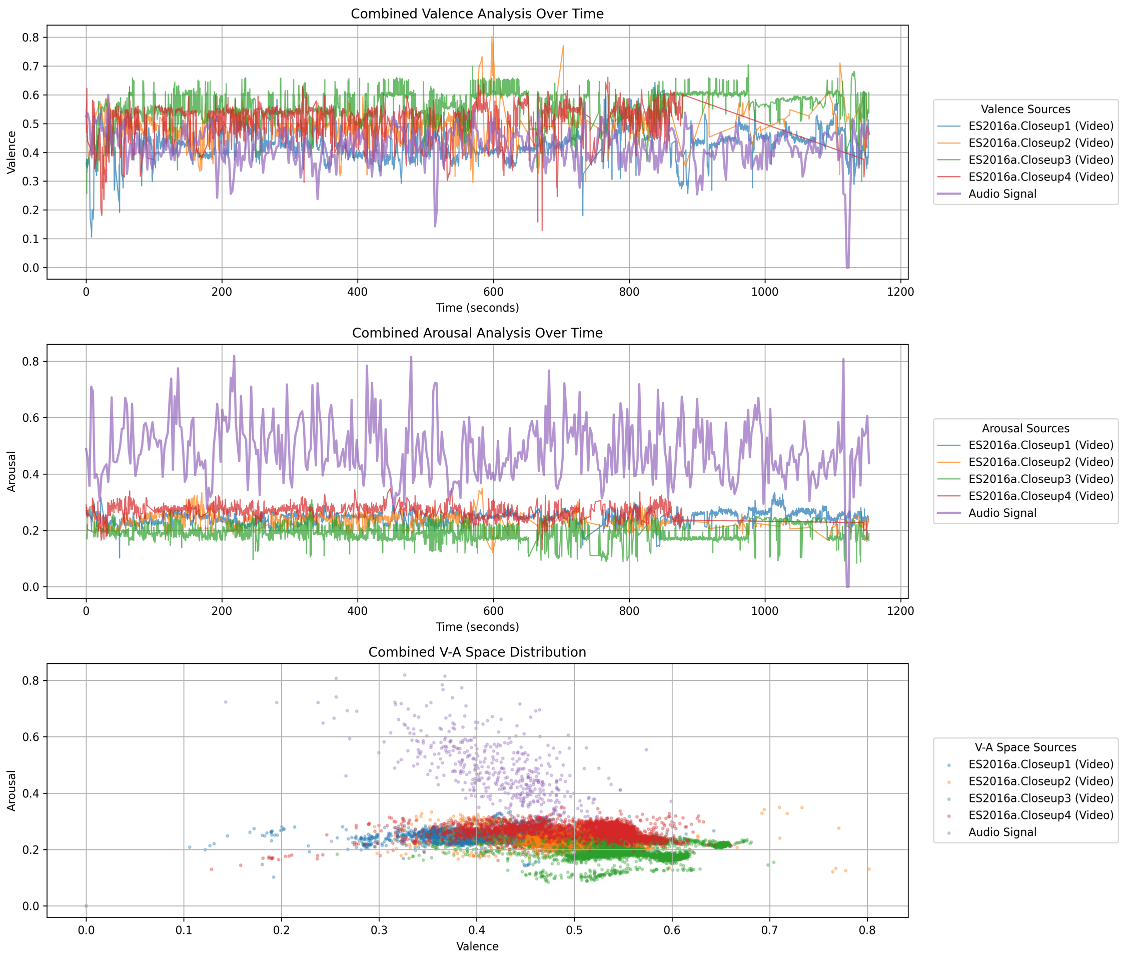
Metric	Threshold	Explanation
High Valence	> 0.5	Values above 0.5 indicate positive emotions, with 1.0 being the most positive
High Arousal	> 0.2	Values above 0.2 indicate increased engagement/energy, with 1.0 being maximum arousal
Combined V-A	V > 0.5 & A > 0.2	Moments with both high positive emotion and engagement

Audio Analysis Thresholds:

Metric	Threshold	Explanation
High Valence	> 0.5	Values above 0.5 indicate positive emotions in speech
High Arousal	> 0.5	Values above 0.5 indicate increased speech intensity/energy
Combined V-A	V > 0.5 & A > 0.5	Moments with both positive speech emotion and high intensity

Visualization

The following visualization combines data from all five sources (4 video streams and 1 audio stream) to show the temporal evolution of valence and arousal, as well as their distribution in the V-A space.



Combined Valence Analysis Over Time

Combined Arousal Analysis Over Time

Combined V-A Space Distribution

Notable High Valence Moments

Video Analysis - Notable High Valence Moments:

Source	Frame	Time	Valence	Arousal
ES2016a.Closeup2	33315	1110.50s	0.710	N/A
ES2016a.Closeup2	32810	1093.67s	0.609	0.164
ES2016a.Closeup2	34465	1148.83s	0.608	0.223
ES2016a.Closeup1	31730	1057.67s	0.594	0.264
ES2016a.Closeup3	31730	1057.67s	0.594	0.227

Audio Analysis - Notable High Valence Moments:

Time	Valence	Arousal

39.00s	0.599	N/A

Notable High Arousal Moments

Video Analysis - Notable High Arousal Moments:

Source	Frame	Time	Valence	Arousal
ES2016a.Closeup1	31730	1057.67s	0.594	0.264
ES2016a.Closeup2	34155	1138.50s	N/A	0.253
ES2016a.Closeup3	31730	1057.67s	0.594	0.227
ES2016a.Closeup2	34465	1148.83s	0.608	0.223
ES2016a.Closeup2	32810	1093.67s	0.609	0.164

Audio Analysis - Notable High Arousal Moments:

Time	Valence	Arousal
261.00s	N/A	0.819

Notable High Valence AND Arousal Moments

The following moments represent peaks in both valence and arousal, indicating instances of high positive emotional intensity. These moments are particularly significant as they capture periods where participants showed both strong positive emotions and high engagement.

Video Analysis - Combined V-A Peaks:

Source	Frame	Time	Valence	Arousal
ES2016a.Closeup1	31730	1057.67s	0.594	0.264
ES2016a.Closeup2	34465	1148.83s	0.608	0.223
ES2016a.Closeup3	31730	1057.67s	0.594	0.227

Audio Analysis - Combined V-A Peaks:

Time	Valence	Arousal
734.62s	0.574	0.554
55.16s	0.512	0.561

Key Observations

Synchronized Emotional Moments

Frame 31730 (approximately 17 minutes and 38 seconds into the recording) appears to be a significant moment captured by multiple cameras. Both Closeup1 and Closeup3 show their highest valence (0.594) and arousal (0.264 for Closeup1, 0.227 for Closeup3) at this exact frame, suggesting a synchronized emotional response across multiple participants.

Most Emotionally Varied Camera

Closeup2 shows the most varied emotional moments with multiple peaks. It recorded the highest individual valence (0.710) across all sources, suggesting it captured the most positive emotional moment in the interaction.

Audio-Video Alignment

The audio's highest valence (0.599) is very close to the video's highest (0.594 for Closeup1/3), suggesting alignment between audio and video emotional signals. However, the audio shows much higher arousal peaks (0.819) compared to video (0.268), indicating that audio may be more sensitive to arousal changes or that high-arousal moments may not always be visible in facial expressions.

Conclusion

This multi-modal emotion analysis provides valuable insights into the emotional dynamics of the AMI Corpus ES2016a session. By combining video and audio data, we gain a more comprehensive understanding of the emotional landscape of the interaction. The analysis reveals both synchronized emotional moments across multiple participants and modality-specific patterns in emotional expression.