



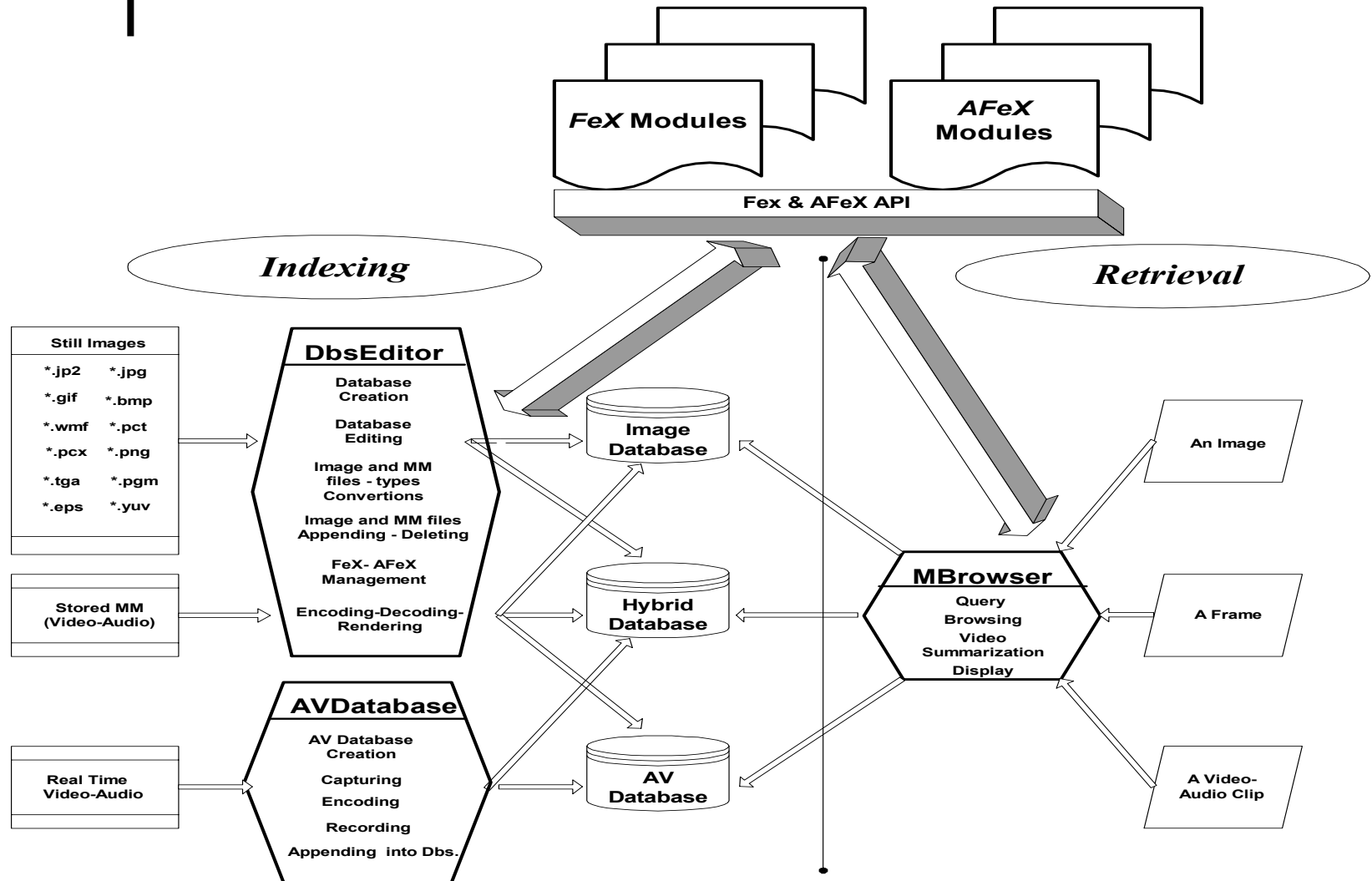
Audio-Based Multimedia Indexing and Retrieval Framework in MUVIS

System Overview & Applications

by Serkan KIRANYAZ.

Tampere Univ. of Tech.

MUVIS Overview





MUVIS Multimedia

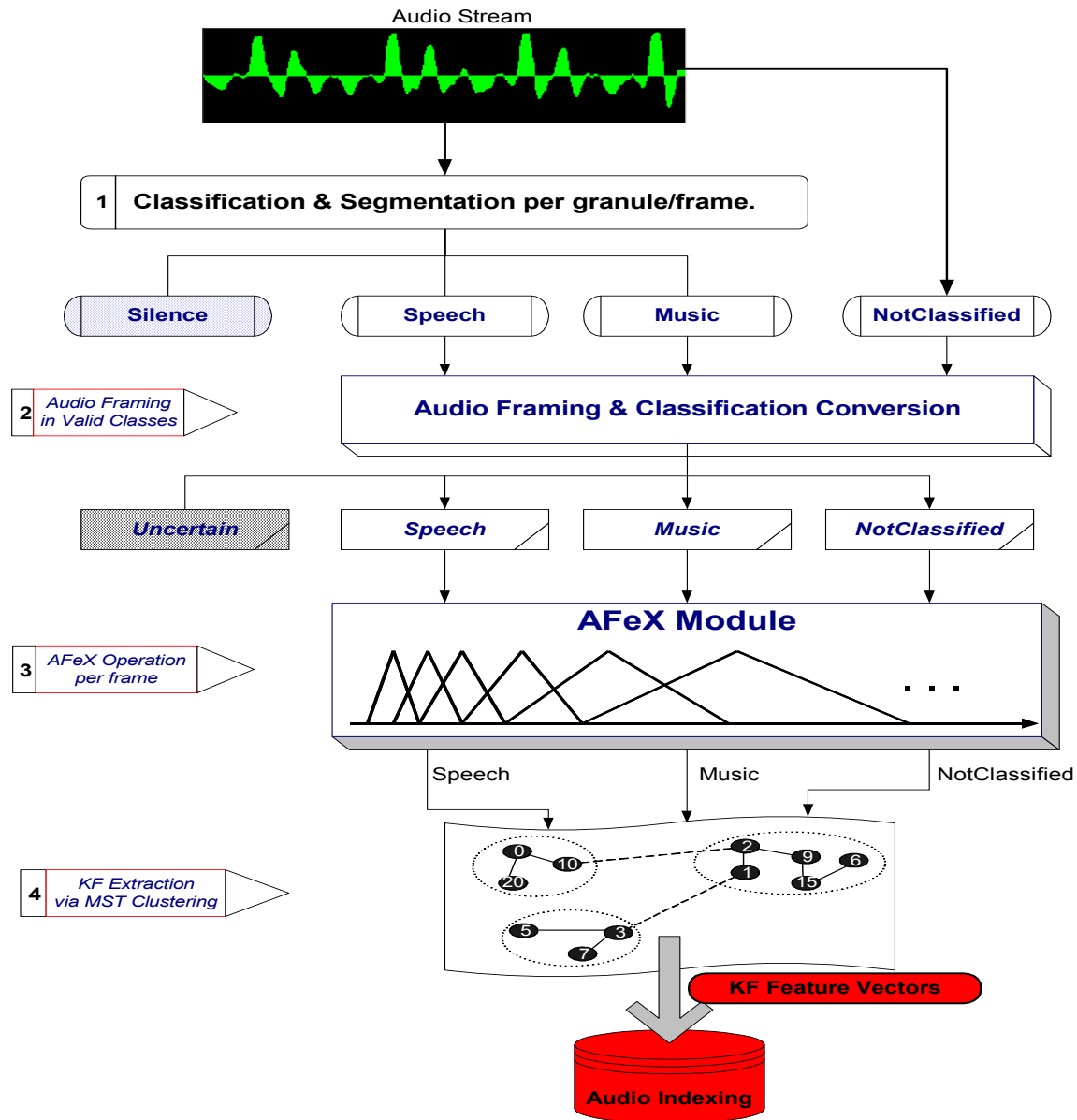
MUVIS Audio				MUVIS Video				
Codecs	Sampling Freq.	Channel	File Formats	Codecs	Frame Rate	FrameSize	File Formats	
MP3	8 & 11.025 Khz	Mono	MP3	H263+	1..25 fps	Any	AVI	
AAC	12 & 16 KHz	Stereo	AAC	MPEG-4			MP4	
G721	22.050 KHz		AVI	YUV 4:2:0				
G723	24 KHz		MP4	RGB 24				
PCM	32 KHz							
	44.1 KHz							

MUVIS Images												
JPEG	JPEG 2K	BMP	TIFF	PNG	PCX	GIF	PCT	TGA	PCX	EPS	WMF	PGM

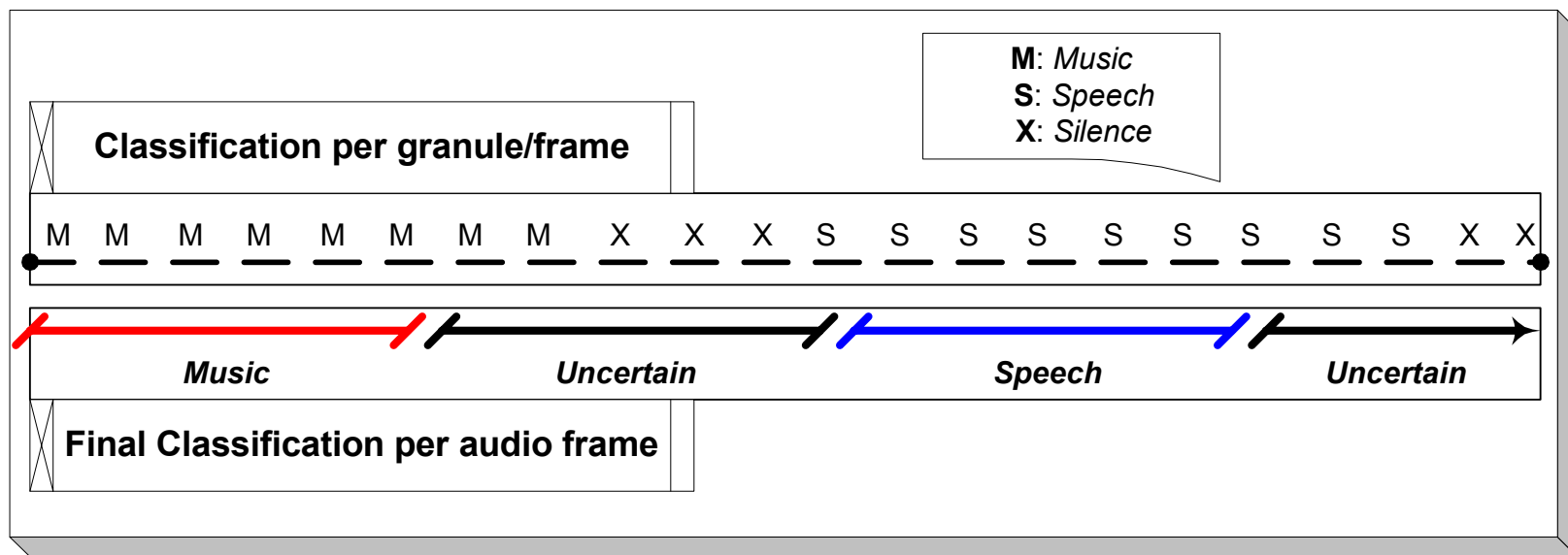
Audio-Based Multimedia Indexing and Retrieval Framework for MUVIS

- A global framework implementation in order to achieve a robust and generic solution for audio-based multimedia indexing and retrieval, specifically:
 - **Generic Support for Audio Codecs**
 - **Generic Support for File Formats**
 - **Generic Support for Audio Capturing & Encoding Parameters**
 - **Generic Support for *AFeX* Framework Parameters**
- The main objective is content-based (speaker, subject, “sounds like..”) retrieval of the audio, which is suitable to human judgment and (aural) perception.

Audio Indexing Scheme in MUVIS

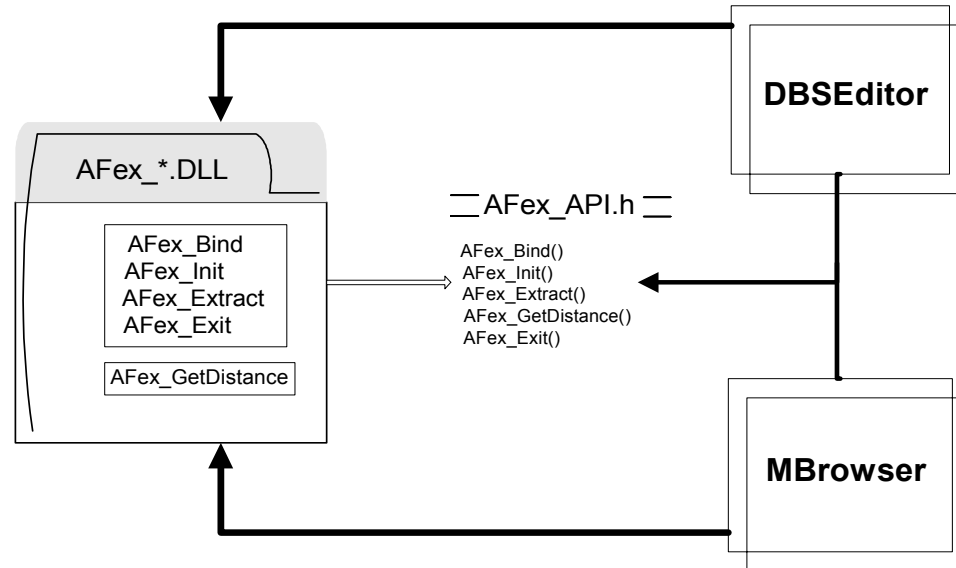


2. Audio Framing with Classification Conversion

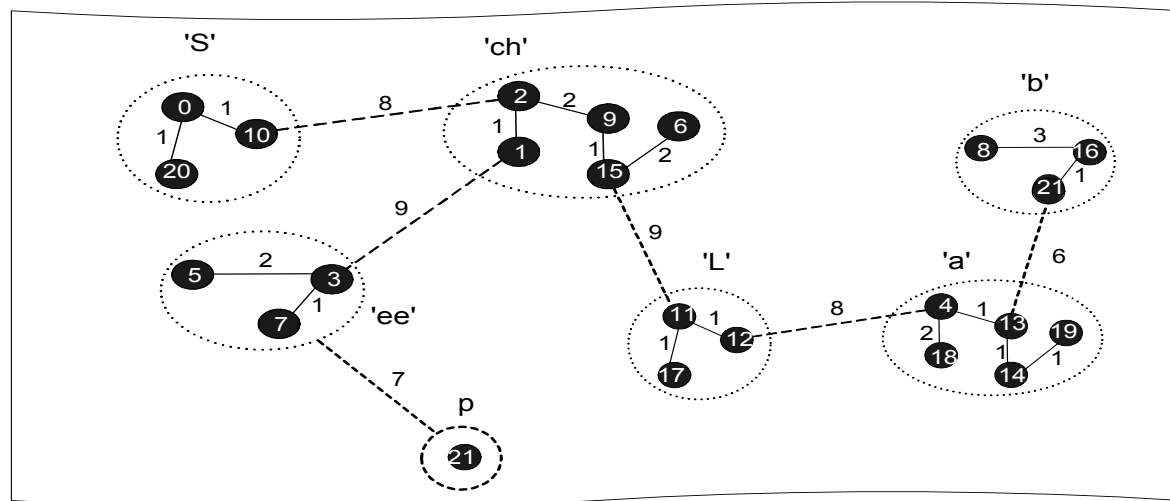
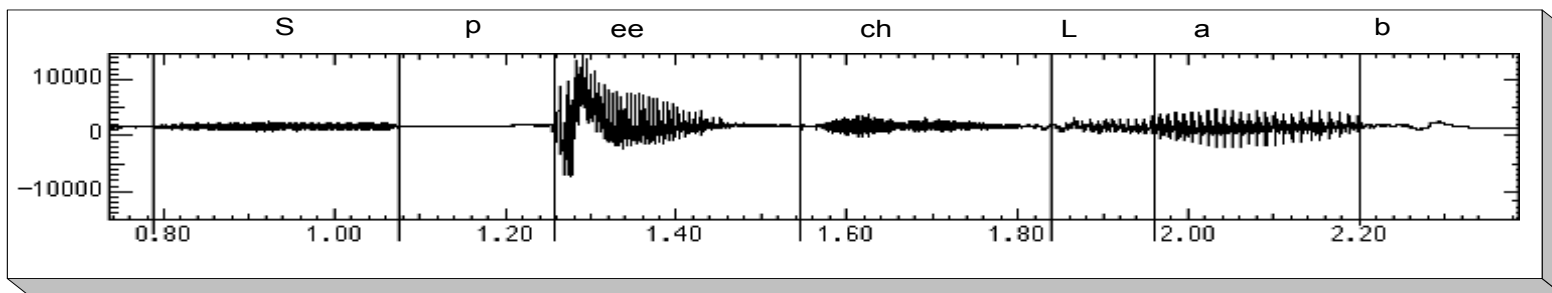


Audio Feature Extraction (AFeX) Framework

- Independent AFeX module(s) integration capability into MUVIS framework for audio-based indexing and retrieval.

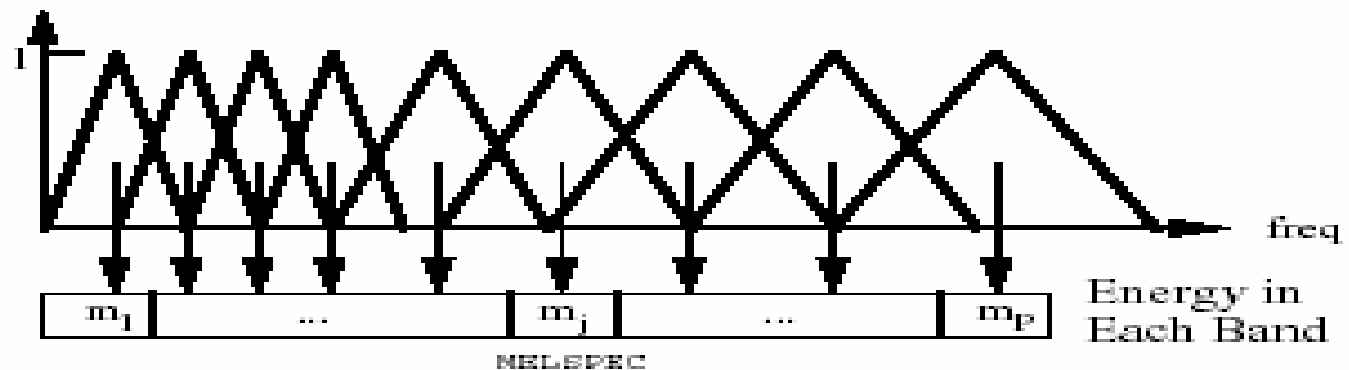


Key-Framing via MST Clustering



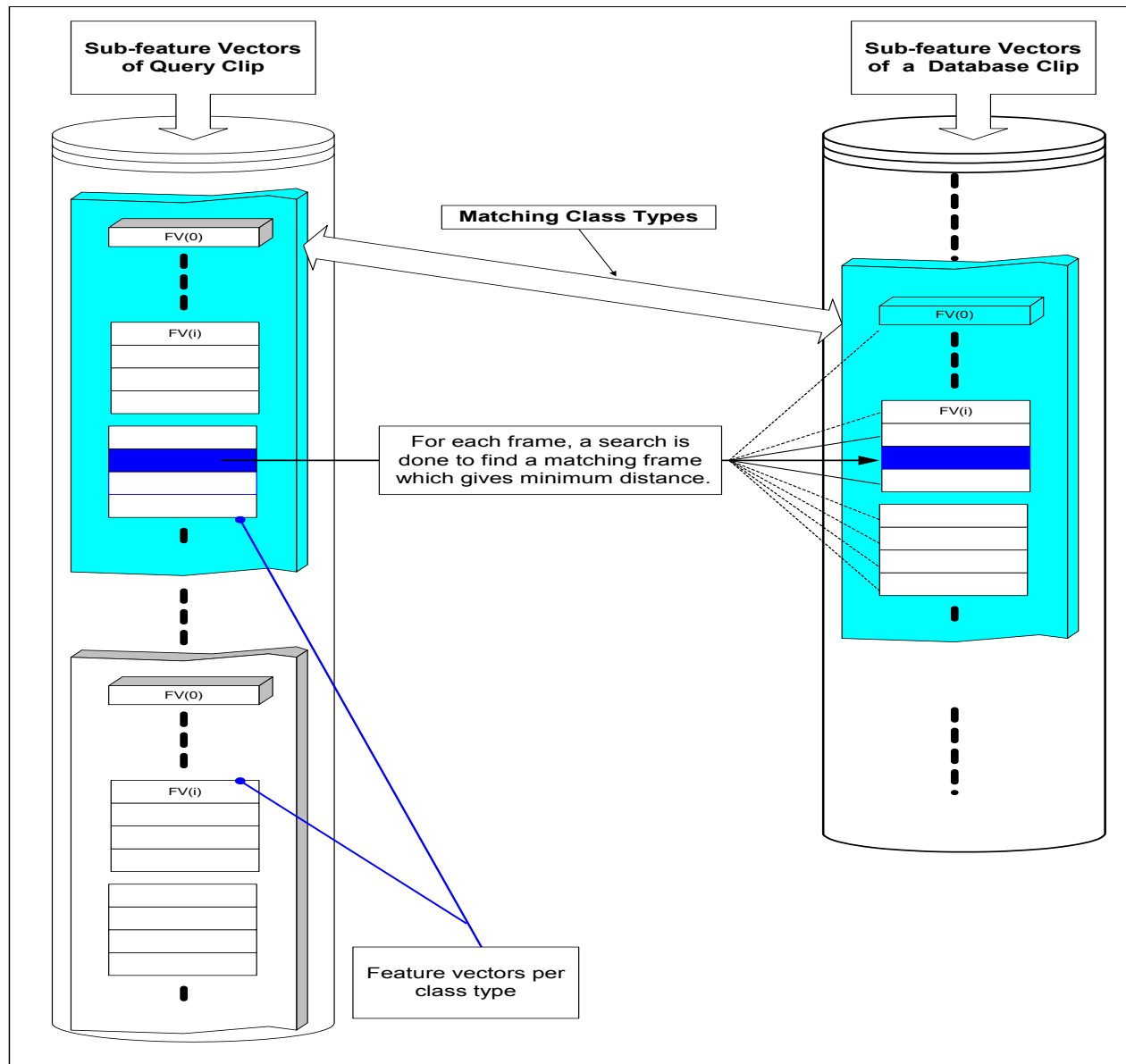
A Sample *AFeX* Module Imp.: MFCC

- MFCC (Mel-Frequency Cepstrum Coefficients) *AFeX* module provide generic feature vectors independent from the following parameters:
 - Sampling Frequency.
 - Number of audio channels (mono/stereo).
 - Audio Volume level.



$$c_i = (2 / P)^{1/2} \sum_{j=1}^P \log m_j \cdot \cos \left(\frac{\pi \cdot i}{N} (j - 0.5) \right)$$

Audio Retrieval in MUVIS



Audio Retrieval in MUVIS (cont.)

- In order to accomplish an audio based query within MUVIS, an audio clip is chosen in a multimedia database and queried through the database if the database includes at least one audio feature.
- Let NoS be the number of feature sets for a database and let $NoF(s)$ is the number of sub-features per feature. Sub-features are obtained by changing the $AFeX$ module parameters or the audio frame size during the audio feature extraction process.

$$D_i(s, f) = \left\{ \begin{array}{l} \min \left(SD \left(QFV_i^{C_i}(s, f), DFV_j^{C_i}(s, f) \right) \right)_{j \in C_i} \text{ if } j \in C_i \neq \emptyset \\ 0 \text{ if } j \in C_i = \emptyset \end{array} \right\}$$

$$D(s, f) = \sum_i^{i \in C_i} D_i(s, f)$$

$$QD_c = \sum_s^{NoS} \sum_f^{NoF(s)} W(s, f) \times D(s, f)$$



Conclusions & Remarks

- Audio is important. Sometimes it bears more semantic and content information than video.
- Henceforth the preliminary results shows the effectiveness of the audio-based retrieval compared to visual retrievals (similar or better results).
- Classification and segmentation algorithm has been recently improved. A new approach based on fuzzy-regions and semantic-rule-based classification with intra segment boundary detection has been developed.