1. INTRODUCTION

Video has become one of the most popular multimedia artifacts used on PCs and the Internet. The last ten years had witnessed the emergence of any kind of video content. In majority of the cases, within a video, the sound holds an important place.

In the same time, certain individuals are deaf and occasionally cannot understand the meanings of such videos because there is not any text transcription available.

Also, people with different languages and accents are not able to undertstand the videos properly.

Therefore, it is necessary to find solutions for the purpose of making these media artifacts accessible and understandable to most people.

2. AIM- To facilitate the users understand the videos properly with the help of subtitles.

3. CONSTRAINTS : Audio Extraction, Speech Recognition, Acoustic Model, Language Model , Subtitle Generation, Sphinx-4,JAVE.

4. PROJECT DESCRIPTION

Three distinct constraints have been defined, namely **audio extraction, speech recognition** and **subtitle generation** . The system should take a video file as input and generate a subtitle file as output.

4.1 Audio Extraction: The audio extraction routine is expected to return a suitable audio format that can be used by the speech recognition module as pertinent material. It must handle a defined list of video and audio formats. It has to verify the file given in input so that it can evaluate the extraction feasibility. The audio track has to be returned in the most reliable format.

4.2 Speech Recognition: The speech recognition routine is the key part of the system. Indeed, it affects directly performance and results evaluation. First, it must get the type (film, music, information, home-made, etc...) of the input file as often as possible. Then, if the type is provided, an appropriate processing method is chosen. Otherwise, the routine uses a default configuration. It must be able to recognize silences so that text delimitations can be established.

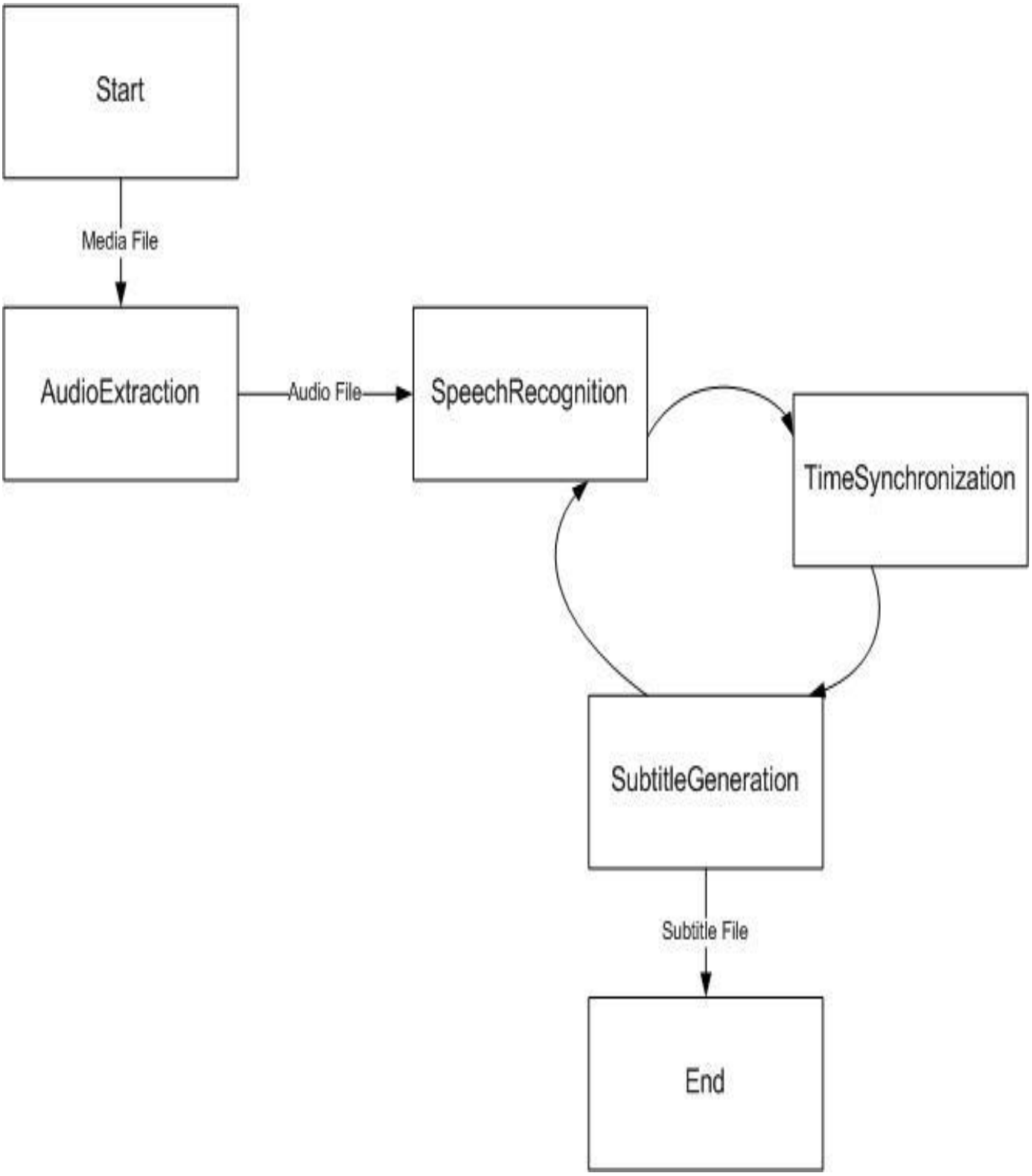
4.3 Subtitle Generation: The subtitle generation routine aims to create and write in a file in order to add multiple chunks of text corresponding to utterances limited by silences and their respective start and end times. Time synchronization considerations are of main importance.

4.4 Acoustic Model: It is a file containing a statistical representation of the distinct sounds that make up each word in the language model. It is supposed to recognize speech from different people. It is obvious a speaker independent acoustic model which requires much more speech audio training to provide correct results.

4.5 Language Model or Grammar: A language model groups a very broad list of words and their probability of occurrence in a given sequence. In a grammar, a list of phonemes is associated to every word. The phonemes correspond to the distinct sounds forming a word.

4.6 Sphinx4: Sphinx4 is a pure Java speech recognition library. It provides a quick andeasy API to convert the speech recordings into text with the help CMUSphinx acoustic models. It can be used on servers and in desktop applications. Beside speech recognition Sphinx4 helps to identify speakers, adapt models, align existing transcription to audio for timestamping and more.

Data Flow Diagram of the experimental system named AutoSubGen.



4.7 JAVE: The **JAVE** (**J**ava **A**udio **V**ideo **E**ncoder) library is Java wrapper on the ffmpeg (Cross-platform solution to record, convert and stream audio and video) project. Developers can take take advantage of JAVE to **transcode audio and video files** from a format to another. JAVE requires a J2SE environment 1.4 or later and a Windows or Linux OS on a i3 86 / 32 bit hardware architecture. JAVE can also be easily ported to other OS and hardware configurations.
In this project, JAVE is used for converting video(.mp4) format to audio(.wav) format.

5 . Software Requirements: Netbeans, Adobe Photoshop

6. Language Requirements: Core and Advanced JAVA

7 . Approach of Development : Evolutionary model :Prototyping

7.1 Details of Methodology:

- Software system evolves over time as requirements often change as development proceeds. Thus, a straight line to a complete end product is not possible. However, a limited version must be delivered to meet competitive pressure.
- Usually a set of core product or system requirements is well understood, but the details and extension have yet to be defined.
- You need a process model that has been explicitly designed to accommodate a product that evolved over time.
- It is iterative that enables you to develop increasingly more complete version of the software.

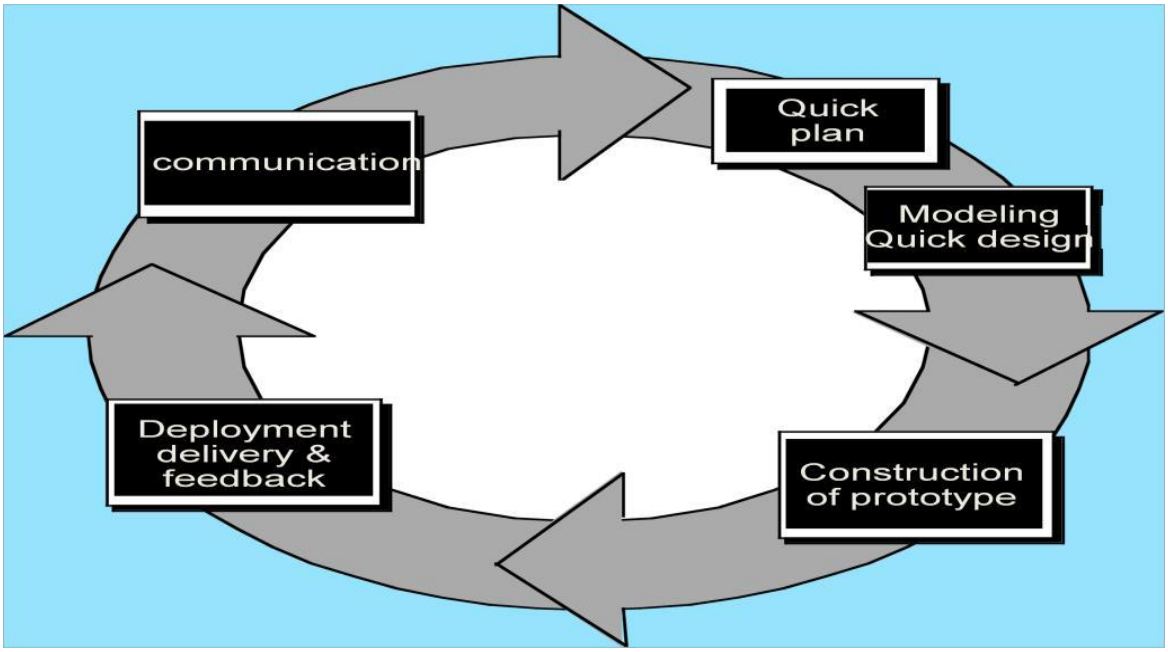


Figure2: Evolutionary model:Prototyping

8. Project Scope: In a majority of cases within a video, the sound holds an important place. From this statement, it appears essential to make the understanding of a sound in video available for people with auditory problems as well as for people with gaps in the spoken language. The most natural way lies in the subtitles.

However, manual subtitle creation is a long and boring activity and requires the presence of the user. Consequently, the study of the automatic subtitle generation appears to be a valid subject for research.

9.REQUIREMENTS:

9.1 Functional Requirements:

- MP4 format is supported for video.
- Audio of **.wav** format is to be converted from video.
- The extracted text from the audio is in the **.srt** format. The text displayed will have a readable format.
- Captions appear on-screen long enough to be read. It is preferable to limit on-screen captions to no more than two lines.
- Captions are synchronized with spoken words.

9.2 Non-Functional Requirements:

- **System Requirements** – The software is compatible on all the Operating Systems. The user needs to install the .exe file of the software in their PCs.
- **Security** – The system has no security constraints.
- **Performance** – The text is synchronized with the song.
- **Maintainability** – The software is easy to maintain.
- **Reliability** - The software will provide a good level of precision.
- **Modifiability**- The software cannot be modified by external user.
- **Scalability-** The software is scalable as a number of users can utilize it for their benefits simultaneously.