UpStage Text-to-Speech & Video Analysis

# Foreword

To ensure we find feasible solutions for the replacement of the current UpStage technologies, here I will discuss the features of Upstage and their existing technologies. This includes both Text to Speech and Video technologies.

As one of the main goals in replacing the current technology is to be usable for mobiles, the replacement of flash is necessary. As discussed in the “UpStage Front-end Analysis” document flash is a majorly dominant technology in UpStage, and is also used in the Text to Speech technology.

# Video Technologies

The current video streaming technology in the UpStage version 3 implementation is unknown. At the beginning of 2014 video was functioning. The video streaming technology used the Red5 server as a buffer for data streams, while the implementation was known to developers as “Martins hack”. However sometime around October 2015 there were various issues with the UpStage public server which no longer allows us to test video streeming. Also since there is little to no documentation on the video streaming technology in UpStage we are unable to determine how this technology was working at the beginning of 2014 nor what technologies it uses.

# How Text-to-Speech is processed

With the help of the proof of concept document created for Text-to-Speech in semester two of 2014, It is determined that the technologies used in UpStage are Festival, MBROLA and eSpeak libraries.

Festival is a multi-lingual speech synthesis API that allows an integration with other speech synthesizers. MBROLA and eSpeak are the speech synthesisers currently used which allow the wide range of voices.

When text is entered into chat these technologies generate an mp3 audio file which is placed in “UpStage2013\server\src\html\speech.py” named “utter-” followed by an incrementing number. The JavaScript file “UpStage2013\server\src\html\script\mediaupload.js” calls “FlowPlayer” to play the audio files in the client’s browser.

# FlowPlayer

FlowplayerFlash is the current technology for UpStage Text-to-Speech that uses streaming technology to play the audio to the user. This technology relies on flash however there is also a Flowplayer html5 version which does not rely on flash, which is both responsive and accessible from mobile devices.

# Solution

The technology that needs replacing for Text-to-Speech is the FlowplayerFlash. To replace this technology will remove any dependencies on flash and allow Text-to-Speech to be accessible by mobile devices. A likely candidate to replace this technology is Flowplayer the html5 version as it has the same amount of developer documentation and a similar implementation style as the existing technologies. Also the Flowplayer technology is focused on streaming technology which would make it a technology suitable for video as well.

There is no issue when it comes to the speech synthesis libraries as they can be used on multiple platforms and therefore will still be usable by the replacement streaming technology.

# Refrences

Archlinux(2015). Festival. *Retrieved from* [*https://wiki.archlinux.org/index.php/Festival*](https://wiki.archlinux.org/index.php/Festival)

Festvox(2014). University of Edinburgh Festival Speech Synthesis System. *Retrieved from* [*http://www.festvox.org/festival/index.html*](http://www.festvox.org/festival/index.html)

MBROLA(2005). The MBROLA Project. *Retrieved from* [*http://tcts.fpms.ac.be/synthesis/mbrola.html*](http://tcts.fpms.ac.be/synthesis/mbrola.html)

Flowplayer. The video player for the web. *Retrieved from* [*https://flowplayer.org/*](https://flowplayer.org/)