

CRIMSONWING IS DEVELOPING THE MALTESE SPEECH ENGINE

Crimsonwing (Malta) Ltd has been awarded the tender to develop a SAPI-compliant Maltese Speech Engine by the Foundation for Information Technology Accessibility (FITA).

This ambitious task is the first of its kind for Crimsonwing. Development within the highly specialized field of Speech Synthesis is a very challenging and rewarding endeavor for an IT company.

Speech Synthesis concerns the artificial production of human speech from textual input. Crimsonwing will be responsible for the development of a Text To Speech (TTS) system which converts Maltese language text into speech, together with a full lexicon of the Maltese Language as a second major deliverable.

Such software tools already exist for a variety of languages and one can also find Speech Synthesizers as open source software. One such example, Festival, is used by numerous researchers around the world. It is available online and in six different English dialects.

Although such tools have been in existence for some time and are continuously being refined, a speech engine for the Maltese language is not yet available. The provision of this type of software will facilitate access to a vast variety of Maltese electronic text, particularly for readers with special needs.

Currently there is an urgent need for such technology. Users have to rely on the available TTS engines for the English language which does not work adequately when used for Maltese Text to Speech.

FITA's survey research shows that Speech Synthesis has diverse markets. However, there are strong indications that one of the main uses of Speech Synthesis would be for people with physical, sensory or cognitive requirements.

James Bonello, Managing Director at Crimsonwing, said "We are proud to have won this tender in February 2010, as this project will make a significant difference to many Maltese speaking people, especially for partially sighted, blind, illiterate, as well as injured and physically disabled people."

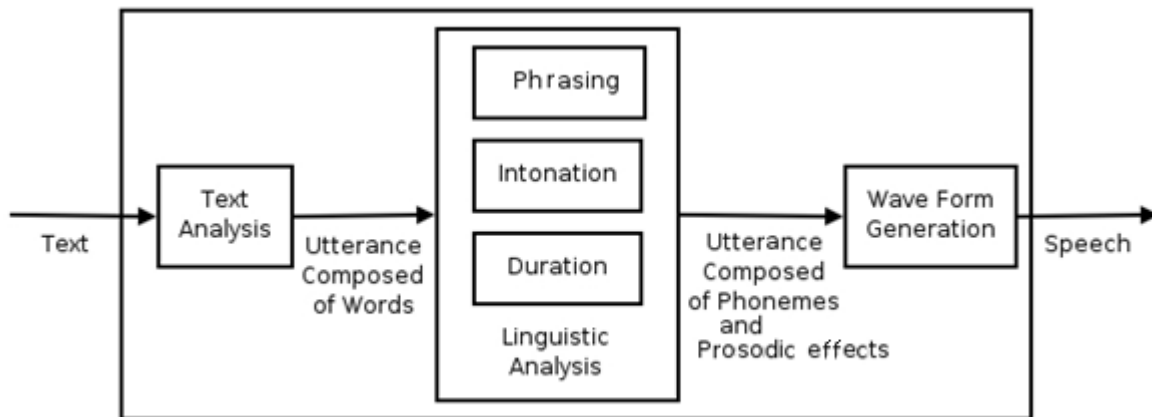
The software would also be capable of reading back students' typed text, for them to hear what they have written and make revisions. It also offers additional speech control features such as tone, pitch, rate of speech and even speaker gender and age. The quality of a speech synthesizer is assessed and judged primarily by its intelligibility and naturalness, apart from other acceptance criteria.



Operational Programme I – Cohesion Policy 2007-2013
Investing in Competitiveness for a Better Quality of Life
Project part-financed by the European Union
European Regional Development Fund (ERDF)
Co-financing rate: 85% EU Funds; 15% National Funds



A typical overview of how a speech synthesizer works is depicted below although many different methods exist for speech synthesis. The system is composed of three major stages. The first stage is responsible for analyzing the text, for instance, by converting numbers and abbreviations into their text equivalent. The second stage performs a linguistic analysis, to determine prosodic effects such as phrasing and intonation. The third and final stage converts the resulting phoneme stream and prosodic effects to an acoustic digital signal.



Picture: Text-to-speech system

The second major deliverable for the FITA project is the development and population of a Maltese Lexicon. This lexicon is an integral part of the software since it is envisaged that the Maltese Text to Speech engine will be a Lexicon-based solution. The Lexicon would be used primarily by the second stage in the system illustrated above, to determine the correct phonetic pronunciation for each word in the input. To handle unknown words, the engine would fall back to a grapheme-to-phoneme algorithm to generate the phonetic pronunciation.

The third stage is envisaged to use unit selection synthesis to generate the acoustic speech signal. This technology entails matching the incoming phoneme stream and prosodic effects to corresponding pre-recorded sound segments from a unit database, that are eventually concatenated to yield the acoustic signal.

The unit database requires a moderately large, recorded Maltese speech transcript from which a broad range of sound units must be extracted. To facilitate concatenation, these units are typically extracted in sound pairs, or diphones. These units must not only cover the complete spectrum of such sound combinations, but must also come in sufficient variety to reflect a diverse range of prosodic effects, such as different stresses or voice pitch formant frequencies.

To priorities unit extraction, statistical analysis will be performed on both written and recorded text transcripts in order to determine, for instance, which diphone combinations require the most



variation. Whilst this is a very laborious process, this variety is the key to achieving natural-sounding artificial speech.

Further to this, the tender mandates the delivery of three voices: an adult male, adult female and a child. This means that speech recording and diphone extraction must be performed for each of the three voices. Most of the above work cannot be done entirely manually and Crimsonwing is already planning to develop software tools which will facilitate the complete statistical analysis and build-up of the Lexicon, Prosody and Unit Databases.



Crimsonwing is investing a lot of effort into research and has to come up with already proven, yet state of the art methodologies to build up the Maltese Speech Engine and the Maltese Lexicon.

An indispensable requirement for any project, let alone a complex one, is a detailed project plan. This has been drawn as one of the initial processes when the tender was awarded. Work schedules and timings are continuously being assessed by FITA and Crimsonwing.

As part of the above mentioned plan Crimsonwing is currently focusing on major research areas, namely:

- Concatenative Speech Synthesis Methodologies
- Prosody Modeling
- Unit Selection Synthesis

This is the most complex part of the research area, which covers the build up and development of the algorithmic work required for the speech engine, that is, the logic of how the sound units will ultimately be selected in relation to the textual input.

SAPI COMPLIANCE

An important requirement for this project is that the Speech Engine will have to be compatible with SAPI. SAPI stands for Speech Application Programming Interface, which is an API defined by Microsoft for Speech Recognition and Speech Synthesis. This API specifies a standard interface that guarantees interoperability with the majority of speech-enabled applications and assistive technologies on the Microsoft Windows platform. Typical examples of such applications are Window Eyes and Zoom Text.



Operational Programme I – Cohesion Policy 2007-2013
Investing in Competitiveness for a Better Quality of Life
Project part-financed by the European Union
European Regional Development Fund (ERDF)
Co-financing rate: 85% EU Funds; 15% National Funds



FRAMEWORK DESIGN

This part of the research is the analysis on integration methods for all the tools which need to be developed and which will assist Crimsonwing in completing the speech engine, lexicon and unit databases. These tools will be complex and not necessarily written in the same development computer language so proper research needs to be conducted to harmonize these tools which will need to interact with each other.

FIRST GENERATION PROTOTYPE DELIVERY

Crimsonwing resources are also researching the key elements required for the first prototype engine as requested by FITA. This prototype will be based on second generation speech synthesis technology and will also be compliant with the SAPI interface to enable initial user trials.

Crimsonwing has agreed with FITA to deliver the software in a phased approach.

The delivery will consist of a series of prototype deliveries with the first engine prototype earmarked to be delivered in August 2010.

This delivery is very important for both Crimsonwing and FITA. It will ensure that Crimsonwing has a solid proof of concept for their speech engine. It will also ensure that this engine is SAPI-compliant and this will be trialed by all stakeholders, including FITA and the ACTU unit within the Department of Education.

From FITA's perspective, the delivery of the first prototype will be a tangible milestone providing a functional prototype of the Maltese Speech engine that is also compliant with SAPI.

In order to achieve such results, Crimsonwing has invested heavily in its human resources. It has assigned five technical architects having several years of experience within Crimsonwing. In addition, technically experienced Crimsonwing management personnel are directly involved in the steering of the project with FITA.

Crimsonwing is working closely with leading consultants from the University of Malta as an integral part of the development team. These experts come from different areas of specialization, namely Engineering, Linguistics and Speech Therapy.

This project is co-financed via an 85% grant by the European Regional Development Fund and 15% by the Maltese government. Being EU funded, the project falls under the EU's cohesion policy 2007-2013 "Investing in Competitiveness for a Better Quality Of Life".



Operational Programme I – Cohesion Policy 2007-2013
Investing in Competitiveness for a Better Quality of Life
Project part-financed by the European Union
European Regional Development Fund (ERDF)
Co-financing rate: 85% EU Funds; 15% National Funds

