"Localization of speech synthesis systems for Irish voice web operations"

R. Hogan, J. Timoney
Department of Computer Science,
NUI Maynooth,
Maynooth,
Co. Kildare,
Ireland

R. Villing, T. Ward
Department of Electronic Engineering,
NUI Maynooth,
Maynooth,
Co. Kildare,
Ireland

I. Robertson, T.Ó Dúshláine, R. Ó hUiginn
Department of Nua Ghaeilge,
NUI Maynooth,
Maynooth,
Co. Kildare,

Areas: Services and Applications; Ubiquitous Computing

As the information age creates growing consumer demand for information, anytime and anyplace voice applications and voice based services are dramatically changing the way people and businesses communicate. Voice portals established an early market but the development of advanced voice applications has introduced speech technology to the masses. Retail banking is currently and is expected to remain the largest market for voice applications through 2007 [1]. A key opportunity being explored by Telecommunications Providers (e.g. AT&T) and researchers is the integration of voice applications such as automatic language translation as a network service. At the heart of all such systems are speech recognition (SR) and text-to-speech (TTS) components. The SR is used for more natural user input while the TTS parses the text returned by the web servers and renders it as audible speech through a process of synthetic speech generation. Currently there are no available TTS systems localised for Ireland due to the small market size. While it is possible to build an Hiberno-English accent TTS system suitable for such applications in a short space of time (6-8 weeks) the implementation of Irish Gaelic synthesis is a much more involved problem due to the phonetic differences with English.
In this poster we report on work done to develop both a Hiberno-English and an Irish Gaelic speech synthesis system suitable for such applications. Such synthesis engines are an essential component of any voice application or service. In this work a diphone synthesis model was utilised through the Edinburgh Speech Tools and Festival synthesis systems [2].
To construct the Irish Gaelic speech synthesizer, three essential tasks were recognized: Firstly it was necessary to identify a complete phone set for Irish. Once done it was necessary to create the definition of

the phonetic features for original phones determined to exist in Irish Gaelic. The second task was to convert this phone set into a series of recorded diphones (phone-phone transitions) that would be stored in the system database. The recording procedure used to create a new synthesizer was to generate and record a list of nonsense words containing target diphones. The most effective method to achieve this was to map all Irish phones onto phones in an existing synthesized language (English). This facilitated the automatic generation of nonsense words and also the prompt waveforms are then used as a blueprint for the automatic labeling of the recorded waveforms. The labels were then selected at the occurrence of each diphone and its boundaries marked in the audio stream. The Hiberno-English synthesis was achieved in a similar manner although the Festival system already possessed the requisite diphones for training.

Text-to-phoneme conversion was achieved through the provision of a lexicon look-up file.

This is the first time a complete Irish Gaelic synthesiser has been developed and although subsequent work will focus on the refinement of the synthesis system for Irish voices and languages it is hoped that separately a demonstration of voice portal applications using our system can be realised over the next year.

References
[1] www.mindbranch.com, "Telematics and Automotive Communications" Market Report, No. R415-0098, February 2003.

[2] Black, A. and Taylor, P., 'The Festival Speech Synthesis System', System documentation, Edition 1.4, for Festival Version 1.4.0, (http://www.cstr.ed.ac.uk/projects/festival/manual/festival_toc.html) 17th June 1999