Developing further speech recognition resources for Welsh

Sarah Cooper
Language Technologies Unit
Canolfan Bedwyr
Bangor University
s.cooper@bangor.ac.uk

Dewi Bryn Jones
Language Technologies Unit
Canolfan Bedwyr
Bangor University
d.b.jones@bangor.ac.uk

Delyth Prys
Language Technologies Unit
Canolfan Bedwyr
Bangor University
d.prys@bangor.ac.uk

Abstract

This paper reports on ongoing research into developing large-vocabulary continuous speech recognition (LVCSR) for the Welsh language. We address data design issues and the method for data collection using a purposely designed application for mobile devices. We also discuss the application of the data including the design and collection of a small speech corpus to cover the commands used to control a robotic arm in Welsh on a Raspberry Pi computer the licensing of the project and our hopes for the application of the project resources to other languages.

1 Introduction

This paper presents an overview of the GALLU (Gwaith Adnabod Lleferydd Uwch- IPA: [gaɬɨ], translation: further speech recognition work) project to develop speech recognition technology for the Welsh language. Wales has a population of around 3 million people, of whom around 20% speak Welsh (Office for National Statistics, 2012). Lesser-resourced languages typically lag in digital innovation, including in language technologies. However since 2012, the Welsh Government has updated and revised a strategy for supporting Welsh-language technology. Emphasis is placed on "more tools and resources … to facilitate the use of Welsh, including in the digital environment" (Welsh Government, 2012: 45) and "the development of new Welsh-language software applications and digital services" (Welsh Government, 2013; 12). With funding from the Welsh Government and S4C (the Welsh language television channel), the GALLU project aims to develop speech recognition technology for the Welsh language. The resources will be available under a permissive open-source licence, and will therefore be available for use in a broad spectrum of platforms and devices, including voice control for smart televisions.

2 Previous speech technology for Welsh

Prior to the GALLU project, the most substantial work on Welsh speech technology was developed under the WISPR (Welsh and Irish Speech Processing Resources) project (Prys et al., 2004). Previous work on a diphone-based synthesiser (Williams, 1994; 1995) and also a small speech database for Welsh (Williams, 1999) was built upon by the WISPR project. An improved synthetic Welsh voice was developed as part of the WISPR project as well as an MSAPI interface to Festival for use in Microsoft Windows environments (Bangor University Text to Speech, [no date]). Following the release of the WISPR resources under an open-source (BSD) licence, further work was facilitated to develop commercial Welsh voices by the Language Technologies Unit at Bangor University, by the Finnish company Bitlips (Bitlips Text to Speech, [no date]) and the Polish company Ivona (Ivona Text to Speech, [no date]). A "Basic Welsh speech recognition" (Bangor University, [no date]) project at the Language Technologies Unit at Bangor University in 2008-9 resulted in laboratory prototypes for a) a "command and control" application for a PC where applications could be launched by voice control and b) a simple voice-driven calculator. The GALLU project will build on this to develop further Welsh speech recognition resources.

3 Data design

The Welsh language has up to 29 consonants and a large number of vowels: up to 13 monophthongs and 13 diphthongs dependent on the variety (Awbery, 1984; Ball, 1984; Jones, 1984; Ball and Williams, 2001; Mayr and Davies, 2011; amongst others). In order to collect the appropriate data to train an acoustic model within HTK ([no date]), a set of phonetically rich words has been designed for contributors to read aloud. In designing the prompt set it was important to ensure that a small number of prompts contain representations of all of the phonemes in the language. The WISPR project's letter-to-sound rules were rewritten based on data mining from a lexicon, and a list of the most common sounds and words was extracted from a text corpus. The final prompt set will contain approximately 200 prompts (8 words per prompt) covering all of the phonemes in the language which may be recorded by contributors across different sessions.

```
{"identifier": "sample1", "text": u"lleuad, melyn, aelodau, siarad, ffordd, ymlaen, cefnogaeth, Helen"},
{"identifier": "sample2", "text": u"gwraig, oren, diwrnod, gwaith, mewn, eisteddfod, disgownt, iddo"},
{"identifier": "sample3", "text": u"oherwydd, Elliw, awdurdod, blynyddoedd, gwlad, tywysog, llyw, uwch"},
{"identifier": "sample4", "text": u"rhybuddio, Elen, uwchraddio, hwnnw, beic, Cymru, rhoi, aelod"},
{"identifier": "sample5", "text": u"rhai, steroid, cefnogaeth, felen, cau, garej, angau, ymhlith"},
{"identifier": "sample6", "text": u"gwneud, iawn, un, dweud, llais, wedi, gyda, llyn"},
{"identifier": "sample7", "text": u"lliw, yng Nghymru, gwneud, rownd, ychydig, wy, yn, llaes"},
{"identifier": "sample8", "text": u"hyn, newyddion, ar, roedd, pan, llun, melin, sychu"},
{"identifier": "sample9", "text": u"ychydig, glin, wrth, Huw, at, nhw, bod, bydd"}
```

Example 1: Display prompts within the Paldaruo application

A large pronunciation lexicon will be developed and used for speech recognition. The next steps for the project involve further data collection and linguistic model development.

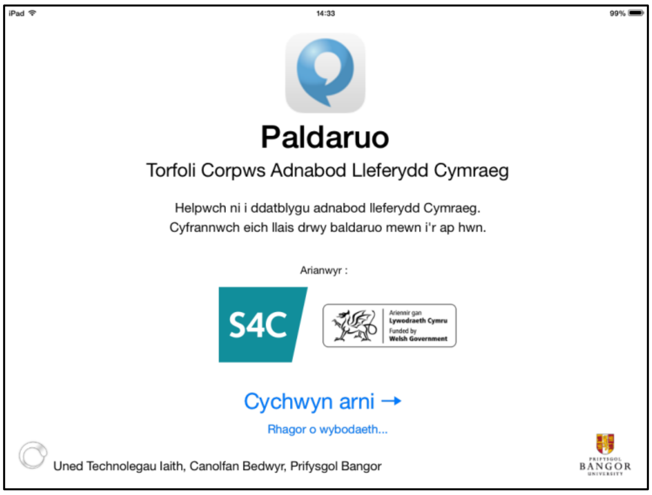
4 Data collection: crowdsourcing and the Paldaruo Application

A large number of speakers are required in order to train the acoustic model which forms the basis of the speech recognition system. Recruiting speakers to attend a recording session at a sound booth with recording software can prove expensive and time consuming. In attempting to tackle this issue, a crowdsourcing approach is being used as a method for collecting data. Crowdsourcing is a low-cost and efficient way of collecting speech data.

A mobile application "Paldaruo" (Welsh for "chattering") has been developed for iOS and Android devices. Such devices, with inbuilt microphones and internet connectivity, provide a convenient mechanism for many volunteers to contribute speech corpus data. The app is optimised for ease of use in order to maximise potential contributions.

Each volunteer creates their own profile within the app providing metadata related to sex, age, linguistic characteristics and geographical background. Following this, the volunteers explicitly agree to their contributions being collected and used. The prompts, described above, are presented one at a time and the volunteer records each one individually. The recording is replayed and the volunteer verifies the quality or re-records. The user can stop and resume at any time. Prompts are provided to the volunteer in a random order; completed prompts will be included in the corpus even if the user does not record the full set.

The app accesses the microphone of the user's mobile device and records 48 kHz PCM files, which are sent to a server developed and hosted by the Language Technologies Unit at Bangor University. Uploads are queued in the background so that network speed issues do not interrupt the recording process.


iPad
14:33
99%
S4C
Ariennir gan Llywodraeth Cymru
Funded by Welsh Government
Prifysgol Bangor University

Paldaruo

Torfoli Corpws Adnabod Lleferydd Cymraeg

Helpwch ni i ddatblygu adnabod lleferydd Cymraeg.
Cyfrannwch eich llais drwy baldaruo mewn i'r ap hwn.

Arianwyr :

Cychwyn arni →

Rhagor o wybodaeth...

Uned Technolegau Iaith, Canolfan Bedwyr, Prifysgol Bangor

Translation:
Paldaruo
Crowdsourcing a Welsh speech recognition corpus

Help us to develop Welsh speech recognition.
Contribute your voice through nattering into this app.

Funders:
S4C Welsh Government

Begin >
More information…

Figure 1: Welcome screen in the Paldaruo App

The app was evaluated in a pilot application (see 5) and proved successful in obtaining a useful speech corpus from invited volunteers. However issues were highlighted with regards to background noise and recording volume levels. To address this, the app now includes background noise and volume level checks.

The official media launch of the app, with the final prompt set, will take place on 7th July 2014. There will be television coverage on S4C with high-profile individuals including the First Minister of Wales and celebrities providing endorsements and appeals for volunteers.

5 Data Application

5.1 Pilot Data Application

To date a small pilot speech corpus has been collected with the Paldaruo app covering the phonemes that appear in a vocabulary to control a robotic arm. 20 speakers contributed to this corpus and recorded 38 prompts (approx. 200 words) each, totalling around 4000 words. Certain commands, for instance ‘up’, exhibit dialect-dependent lexical variation, and in these cases every speaker recorded both regional forms.

Command:	Translation:
golau ymlaen	light on
gafael agor	grip open
gafael cau	grip close
arddwrn i fyny / arddwrn lan	wrist up
arddwrn i lawr	wrist down
penelin i fyny / penelin lan	elbow up
penelin i lawr	elbow down
ysgwydd i fyny / ysgwydd lan	shoulder up
ysgwydd i lawr	shoulder down
troi i’r dde	turn to the right
troi i’r chwith	turn to the left

This corpus has been used to develop a pilot speaker-independent Welsh-language speech recognition system for controlling the robotic arm. The pilot system uses HTK ([no date]) and Julius ([no date]), and follows the design of an existing English system (AonSquared, [no date]). It controls the robotic arm from a Raspberry Pi (a credit card-sized computer, popular in schools and coding clubs, costing around €35; see (Raspberry Pi Foundation, [No date]). The authors hope this simple demonstration will be recreated in schools and coding clubs for children throughout Wales, fitting in with the Welsh

Government's aim to support initiatives aimed at encouraging and supporting young people to engage "in the digital world in a Welsh-language context" (Welsh Government, 2013: 14).

5.2 Licensing

GALLU will follow the successful strategy of the WISPR project in using permissive open-source licensing. All outputs will be made available under the MIT licence (MIT, [No date]) which allows royalty-free use in both open-source and proprietary systems, including desktop computer software, web applications, mobile apps and embedded systems such as television set firmware.

This strategy allows the widest possible use of the project's outputs, and the maximal availability of Welsh speech recognition technology.

5.3 Application to other languages

The authors hope other lesser-resourced languages can harness the project's outputs and experience. The source code of the Paldaruo crowdsourcing app can easily be adapted for use in other languages. The process for developing the LVCSR system has been documented and will be published in the form of a tutorial. All project outputs, including the source code for the app, will be available under the MIT licence.

References

AonSquared. [No date]. *Speech recognition using the Raspberry Pi* [Online]. Available at: http://aonsquared.co.uk/raspi_voice_control [Accessed: 1 May 2014].

Bangor University [No date]. *Bangor University Basic Welsh Speech Recognition Project* [Online]. Available at: http://www.bangor.ac.uk/canolfanbedwyr/adllefsyl.php.en [Accessed: 1 May 2014].

Bangor University Text to Speech [No date] *Festival Demo voice* [Online]. Available at: http://www.e-gymraeg.org/siarad [Accessed: 1 May 2014].

Bitlips Text to Speeeh. [No date]. *Welsh Text to Speech Demo* [Online]. Available at: bitlips.fi/tts/demo-cy.cgi [Accessed: 1 May 2014].

Briony Williams. 1994. Diphone synthesis for the Welsh language. *Proceedings of the 1994 International Conference on Spoken Language Processing,* Yokohama, Japan: 739-742.

Briony Williams. 1995. Text-to-speech synthesis for Welsh and Welsh English. *Proceedings of Eurospeech 1995*, Madrid, Spain, 2: 1113-1116.

Briony Williams. 1999. A Welsh speech database: preliminary results. *Proceedings of Eurospeech 1999,* Budapest, Hungary, 5: 2283-2286.

Delyth Prys, Briony Williams, Bill Hicks, Dewi Jones, Ailbhe Ní Chasaide, Christer Gobl, Julie Carson-Berndsen, Fred Cummins, Máire Ní Chiosáin, John McKenna, Rónán Scaife and Elaine Uí Dhonnchadha. 2004. *SALTMIL Workshop at LREC 2004: First Steps in Language Documentation for Minority Languages.*

Glyn E. Jones. 1984. The distinctive vowels and consonants of Welsh. In Martin J. Ball and Glyn E. Jones (eds.). *Welsh phonology: Selected readings.* University of Wales Press, Cardiff, UK: 40-64.

Gwenllian M. Awbery. 1984. Phonotactic constraints in Welsh. In Martin J. Ball and Glyn E. Jones (eds.). *Welsh phonology: Selected readings.* University of Wales Press, Cardiff, UK. 65-104.

HTK. [No date]. *Hidden Markov Toolkit* [Online]. Available at: http://htk.eng.cam.ac.uk/ [Accessed: 1 May 2014].

Ivona Text to Speech. [No date]. *Text to Speech Portfolio* [Online]. Available at: http://www.ivona.com/en/ [Accessed: 1 May 2014].

Julius. [No date]. *Open-Source Large Vocabulary CSR Engine Julius* [Online]. Available at: http://julius.sourceforge.jp/en_index.php [Accessed: 1 May 2014].

Martin J. Ball and Briony Williams. 2001. *Welsh phonetics*. The Edwin Mellen Press, Lampeter, UK.

Martin J. Ball. 1984. Phonetics for phonology. In Martin J. Ball and Glyn E. Jones (eds.). *Welsh phonology: Selected readings.* University of Wales Press, Cardiff, UK.

MIT. [No date]. *The MIT License* [Online]. Available at: http://opensource.org/licenses/mit-license.html [Accessed: 1 May 2014].

Office for National Statistics. 2012. 2011 Census: Welsh language profile, unitary authorities in Wales. Available at: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-unitary-authorities-in-wales/rft-table-ks208wa.xls [Accessed: 1 May 2014].

Robert Mayr and Hannah Davies. 2011. A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association,* 41(1): 1-25.

Raspberry Pi Foundation. [No date]. *What is a Raspberry Pi?* [Online]. Available at: http://www.raspberrypi.org/help/what-is-a-raspberry-pi/ [Accessed: 1 May 2014].

Welsh Government. 2012. *A living language, a language for living*. Available at: http://wales.gov.uk/docs/dcells/publications/122902wls201217en.pdf [Accessed: 1 May 2014].

Welsh Government. 2013. *Welsh language Technology and Digital Media Action Plan*. Available at: http://wales.gov.uk/docs/dcells/publications/230513-action-plan-en.pdf [Accessed: 20 June 2014].