PRIFYSGOL
BANGOR
UNIVERSITY

Developing further speech recognition resources for Welsh

Sarah Cooper, Dewi Bryn Jones and Delyth Prys with David Chan

Language Technologies Unit, Canolfan Bedwyr,
Bangor University

Llywodraeth Cymru
Welsh Government

S4C

GALLU

- Gwaith Adnabod Lleferydd Uwch
- “Further speech recognition work”
  - Developing speech recognition technology for the Welsh language

Welsh + technology

- ~ 562,000 speakers of Welsh in Wales
- “small” number of speakers
- Languages with small numbers of speakers under-resourced:
  - Availability of funding
  - Interest in funding

Funders:

Encourage "the development of new Welsh-language software applications and digital services"

(Welsh Government, 2013; 12)

- Smart televisions
- Gaming systems

Aims of the GALLU project

- Collect a new Welsh speech corpus through crowdsourcing
- Develop resources for a large-vocabulary continuous speech recognition system (LVCSR) based on this corpus
- Develop a script to control a toy robot using Welsh speech commands (for the Raspberry Pi)

Previous work on Welsh speech technology

- WISPR project (Welsh and Irish Speech Processing Resources)
  - Resources used by private companies to develop commercial Welsh voices
- Sphinx foundation project on Speech Recognition
  - Resulted in laboratory prototypes
- Need for improved speech recognition for Welsh language

Data design: preparing the corpus

- Aim: to cover the most common sound combinations in the language
- Prompt design
- 29 consonants, 13 monophthongs, 13 diphthongs dependent on variety

  (Awbery, 1984; Ball, 1984; Jones, 1984; Ball and Williams, 2001; Mayr and Davies, 2011 etc)

Letter to sound rules

- LTS Rules for mapping orthography onto pronunciation
- Data mining using the LTS rules
  - Most common sounds and words in a corpus of speech
- Combinations of singular sounds, 2 sounds (di-phones) and 3 sounds (tri-phones)
- These checked for readability

Data design

- The final prompt set contains 43 prompts (i.e. lines of words)
- 8 words per prompt
- = 344 words
- Around half an hour to 45 minutes to record

Data Collection

- Large number of speakers required for speech recognition
- Speaker variation also desired
- Recording sessions can be costly and time consuming

crowdsourcing, *n.*

- The practice of obtaining information or services by soliciting input from a large number of people, typically via the Internet and often without offering compensation.

OED Online, 2014

The Paldaruo App

- Smart phones and tablets commonplace
- Purpose made App for iOS and Android
- Read a script used to train the speech recognition system

Crowdsourcing and the Paldaruo App

- Metadata:
  - Age
  - Sex
  - Childhood living area
  - Current living area
  - Frequency speaking Welsh
  - Categorise the accent

What does Paldaruo do?

- Background info:
  - Create a profile
    - Multiple profiles on one device
  - Collects the metadata
  - Terms and conditions
  - More information (including video)
  - Background noise check

Paldaruo

Torfoli Corpws Adnabod Lleferydd Cymraeg

Dewisiwch broffil neu gliciwch 'Ychwanegu Proffil' i greu un newydd

Dewi

Rhys

dewi

dewi

rygdhf

Ychwanegu Proffil

Cychwyn >

Uned Technolegau Iaith, Canolfan Bedwyr, Prifysgol Bangor

PRIFYSGOL BANGOR UNIVERSITY

What does Paldaruo do?

- Presents prompts
- Start and end recording
- Automatically *listen t*o recording
  - Option to re-record
- Reports sound levels
  - Prevents too quiet or clipped sound files
- Sends to server

iPad 11:33 100%

Paldaruo

Torfoli Corpws Adnabod Lleferydd Cymraeg

Nôl i'r Cychwyn

Helo Rhys!

12 / 12 testun ar ôl

AGOR AGOR CAU CAU AGOR AGOR CAU CAU

Yn recordio....

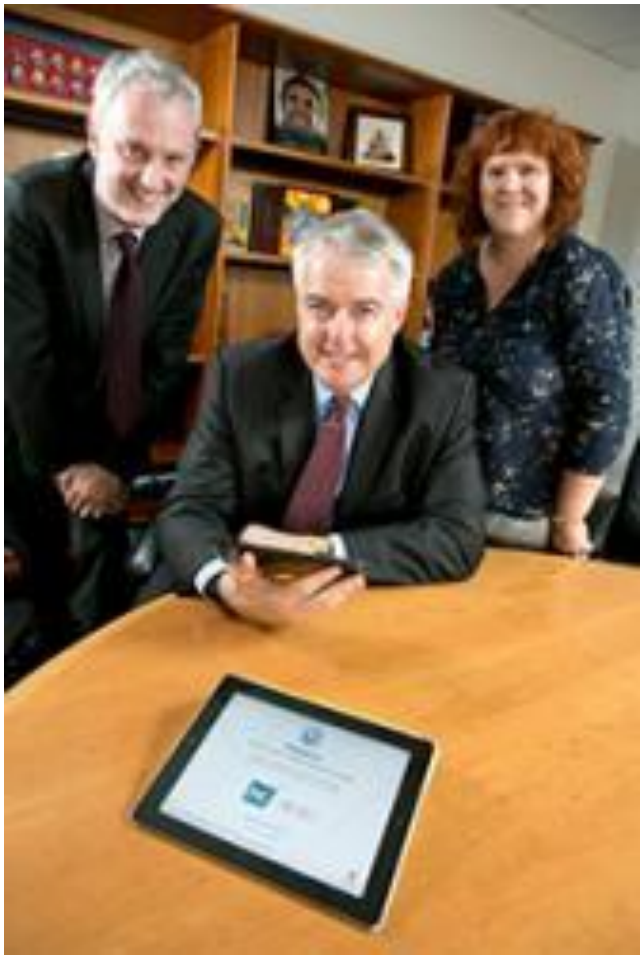
Cliciwch i orffen recordio

Uned Technolegau Iaith, Canolfan Bedwyr

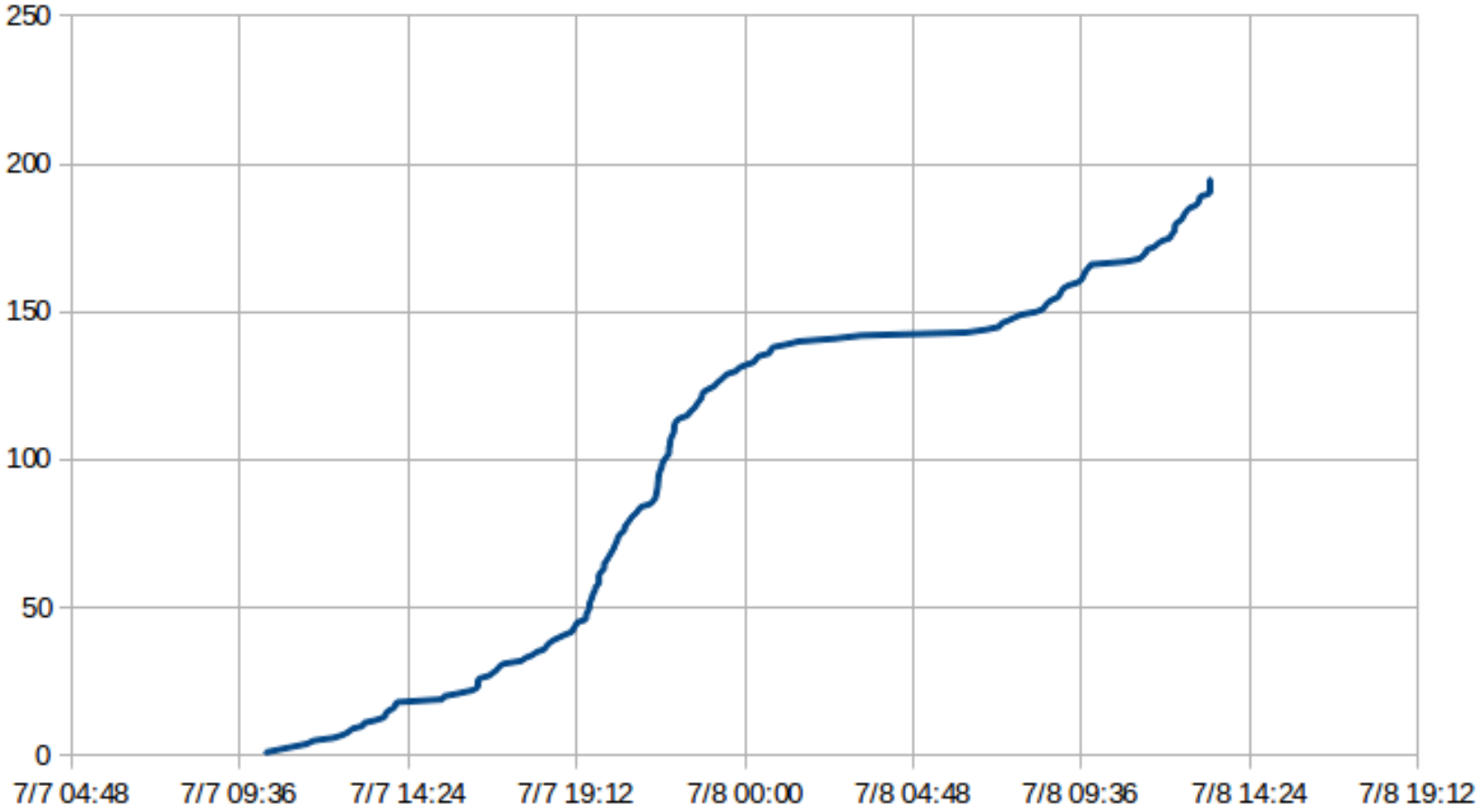
PRIFYSGOL BANGOR UNIVERSITY

Paldaruo Launch (7.7.14)

- Carwyn Jones, First Minister the first to contribute
- Press release and subsequent news stories:
  - Radio:
  - Television
  - News websites
  - Twitter

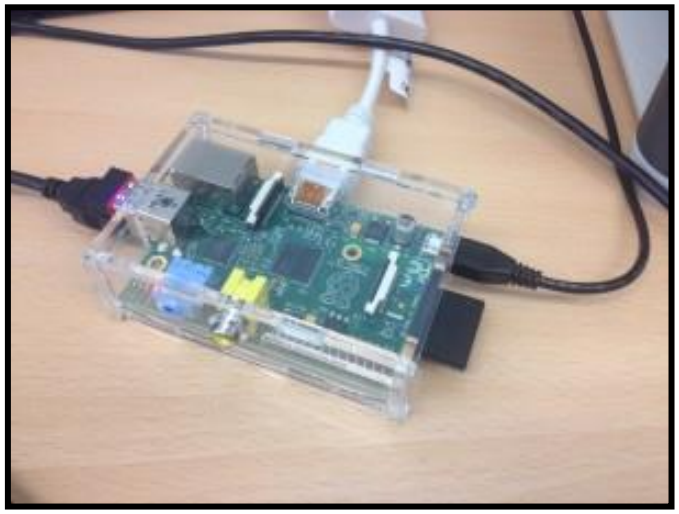


Paldaruo Launch



Pilot Data Application

- Raspberry Pi: small credit card sized computer £30
- Small toy robotic arm
- Use basic speech recognition system to control the robotic arm



Pilot data set

- 20 recordings of people reading prompts designed system for the robotic arm
- Used a beta version of the Paldaruo App to collect the data
- This was useful in raising issues such as noise levels
- Data set available now on http://techiaith.bangor.ac.uk/resources/gallu/samples/

Developing Speech Recogniton

- HTK: Hidden Markov Model Toolkit
  - used primarily for speech recognition research
  - Used to create acoustic models
- Julius: open-source large vocabulary speech recognition engine

Licensing

- permissive open-source licensing
- MIT licence
- allows royalty-free use in both open-source and proprietary systems

Resources available at the end of the project

- Data
- Tutorial on creating acoustic models for robotic arm
- Acoustic models for large data set
- App Source Code

Diolch

techiaith.bangor.ac.uk/gallu

AppStores: Paldaruo

References

Awbery, Gwenllian M. 1984. Phonotactic constraints in Welsh. In Ball & Jones (eds.), 65-104.

Ball, Martin J. & Glyn E. Jones (eds.). 1984. *Welsh phonology: Selected readings.* Cardiff, University of Wales Press.

Ball, Martin J. 1984. Phonetics for phonology. In Ball & Jones (eds.), 5-39.

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

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].

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].

Jones, Glyn E. 1984. The distinctive vowels and consonants of Welsh. In Ball & Jones (eds.), 40-64.

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

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

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

OED Online, 2014. *crowdsourcing, n.* Oxford University Press. www.oed.com.unicat.bangor.ac.uk/view/Entry/376403 (accessed July 17, 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].

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].

Example prompts

rhybuddio, Elen, uwchraddio, hwnnw, beic, Cymru, rhoi, aelod

rh @ b U1 dd IO | E1 l E n | UW ch r A1 dd IO | h W1 nn W
b EI1 c | c @1 m r U | rh OI1 | AE1 l O d

hyn, newyddion, ar, roedd, pan, llun, melin, sychu

h Y1 n | n EW @1 dd IO n | A1 r | r OE1 dd | p A1 n | ll U1 n
m E1 l I n | s @1 ch U