



Natural Speech Technology
Edinburgh – Cambridge – Sheffield

Browsing Oral History

Phil Green, Oscar Saz, Jonathan Kilgour,
George Powell, Trudy Pankhurst Green,
Mark Gales, Pierre Lanchantin, Yanmin Qian



Natural Speech Technology
Edinburgh – Cambridge – Sheffield



ENGLISH
HERITAGE

EPSRC
Pioneering research
and skills

Oral History

- Captures the personal and historical experiences of individuals
- Many collections:
 - English Heritage
 - Heritage Lottery (funded 2000)
 - British Library
- Memories, reminiscences
- Long Interviews
- No central database..
- Utility limited by retrieval problems.. *The Deep Dark Secret of oral history ...that nobody spends much time listening to or watching recorded and collected interview documents'* (Frisch (2008))

Oral History Collections

- Topic-focussed (but topic may be very wide)
- Data back to the 1990s (at least)
- Variable recording quality
- Metadata
 - Typically summary of an interview
 - Occasionally a transcription
- OOV problems .. Use the summary

A photograph of Brodsworth Hall, a large stone mansion with a central tower, set in a lush garden. In the foreground, there is a large, ornate fountain with multiple tiers, surrounded by colorful flower beds. The sky is blue with scattered white clouds.

DUTY CALLS

BRODSWORTH HALL IN TIME OF WAR

A faded, black and white photograph showing a group of people. In the foreground, a man in a military uniform is shaking hands with a woman. Other people are visible in the background, some looking towards the camera.

Oral History recordings .. WWII experiences

- around 25 interviewees
- ~50 hours

Using Speech Technology to Browse Oral History

- ASR Transcription as a browsing tool
- Link Audio, summary and ASR
- Browsing Tool for Duty Calls ('System1')
 - Transcribed with a Sheffield MGB system
 - Speech activity detection based on DNNs
 - 4 independent speaker adapted systems (2 DNN-HMM and 2 DNN-GMM-HMM) trained on 700 hours of BBC broadcasts
 - Lexicon of 50,000 words
 - 4-gram language model and RNN rescoring, trained on 700 million words
- Demo on <http://brodsworthhall.azurewebsites.net>
- 3 interviews .. with permission





Browsing Oral History

brodsworthhall.azurewebsites.net/#/interview/Garnett-Sheila_04-08-2010

Most Visited Customize Links Announcements Free Hotmail MUSE - The University ... Windows Marketplace Windows Media Windows

Browsing Oral History Upload

Search Submit

Cookie policy

Details Summary Manual Transcription Automatic Transcription Search

Summary

Search Summary

SummaryForGarnett-Sheila_04-08-2010.docx

00:12:48	Description of Woodlands	search
00:05:9	Role of Miners' Welfare	search
00:27:7	Humbolt? Family	search
00:08:6	Opening Percy Jackson Grammar School delayed because of outbreak of war	search
00:24:58	How family came to new estate	search
00:09:0	Father's job at the colliery	search



Recognisers for Duty Calls

System 1

- Sheffield MGB system
- Speech activity detection based on DNNs
- 4 independent speaker adapted systems (2 DNN-HMM and 2 DNN-GMM-HMM) trained on 700 hours of BBC broadcasts
- Lexicon of 50,000 words
- 4-gram language model and RNN rescoring, trained on 700 million words

System 2

- SI Hybrid system,
- sequence trained 700hrs trained on MGB data.
- LM data from MGB systems (includes subtitle data).
- IBM KWS

Duty Calls: Trial Corpus

Ground Truth: 8 interviews manually transcribed

	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
Duration	33'39"	1h13'	16'29"	47'54"	1h43'	12'47"	39'42"	21'38"	5h49'
Segments	666	1,965	259	1,151	2,203	335	949	392	7,919
Words	4,605	12,399	1,992	6,647	18,608	2,041	5,843	2,563	54,698

Duty Calls: using the Summaries

7 interviews contain a summary

	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
Words	544	0	663	1,600	891	505	370	853	5,426

A set of relevant named entities (e.g. “Hooton Pagnell”) were manually identified from the summaries, and matched to spoken occurrences in the audio

	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
Entities	41	0	23	23	74	5	29	45	240

Many of the named entities are not in the vocabulary (~40-50%)
When they are, they are underrepresented in the LM training data (either n-gram or RNN)

Duty Calls: finding named entities

Evaluation of WER and Recall of named entities using two different systems

WER over 30%, recall 35%-40% with a high precision (100%-94%)

System 1	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
WER	31.2%	34.0%	19.0%	43.9%	35.2%	46.2%	27.5%	33.6%	34.4%
Recall	24.4%	N/A	73.9%	26.1%	36.5%	20.0%	51.7%	20.0%	35.4%
System 2	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
WER	37.9%	35.8%	24.0%	45.6%	33.9%	51.1%	33.0%	40.4%	36.3%
Recall	24.4%	N/A	73.9%	39.1%	50.0%	0.0%	55.2%	20.0%	40.8%



Duty Calls: adapting the Language Model from the Summary

Lexicon and language model adaptation using the interview summary:

- A new lexicon is created for each summary adding words from the summary
- A new 4-gram is trained interpolating the baseline LM with the summary n-gram
- RNN is fine tuned using the summary

Small WER improvement, but recall increased to >60%

System 1	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
WER	31.6%	N/A	17.3%	42.3%	35.1%	42.2%	28.6%	31.4%	34.1%
Recall	48.8%	N/A	87.0%	52.2%	68.9%	40.0%	69.0%	62.2%	63.8%

Precision 92%

LM Adaptation Across Summaries

Lexicon and language model adaptation using a pool of all summaries:

System 1	Sheila Coates	Jean Covell	Joyce Durdy	Ernest Egginton	David Gilling	Sheila Miles	Jean Paton	Madge Rouse	Total
WER	30.7%	34.8%	16.2%	42.7%	34.3%	42.4%	27.7%	27.7%	33.6%
Recall	51.2%	N/A	91.3%	39.1%	68.9%	20.0%	58.6%	82.2%	65.4%

Precision 92%

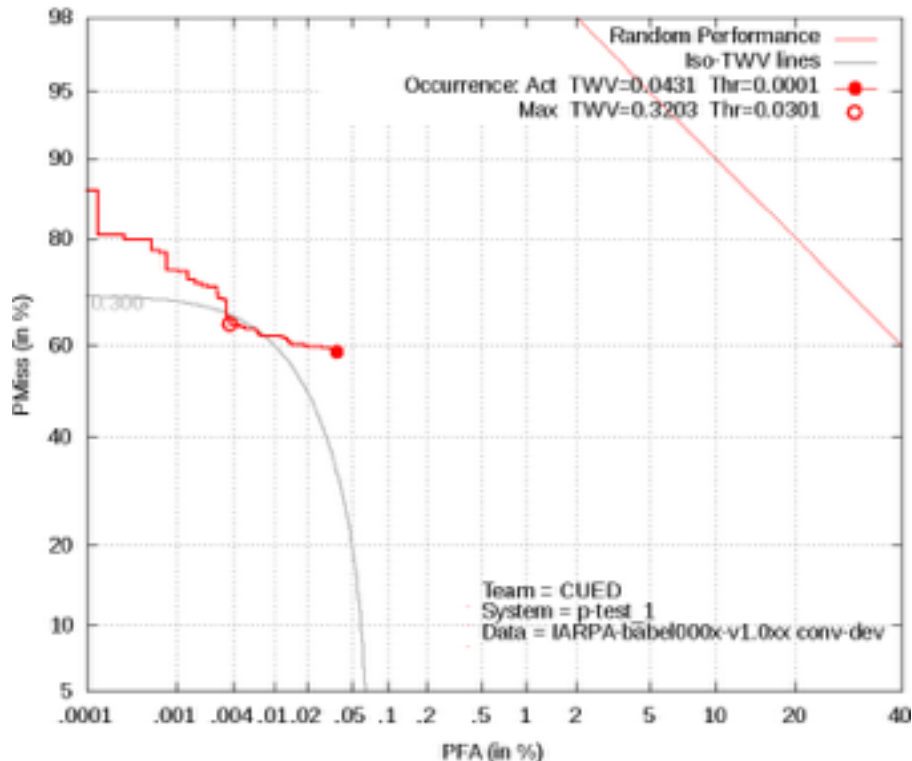


Duty Calls: Keyword Spotting from Lattices

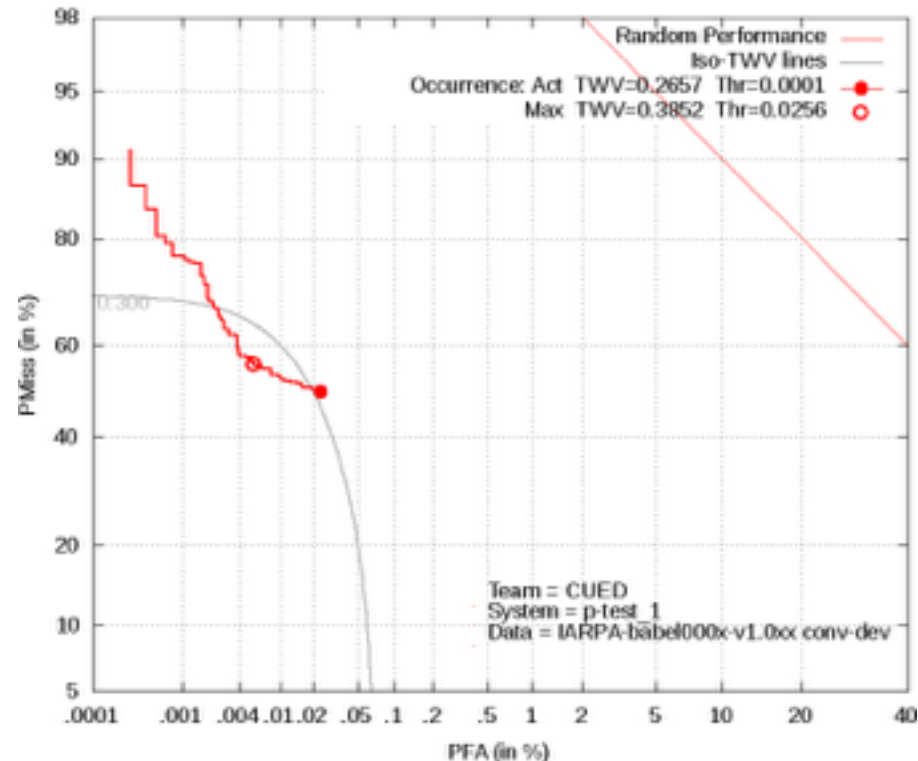
A different operating point can be chosen from the DET curves (keyword spotting based on lattices)

- Recall up to 40-50%

Occurrence scoring, Full Submission, Original orthography, MITLL Force AN. V3



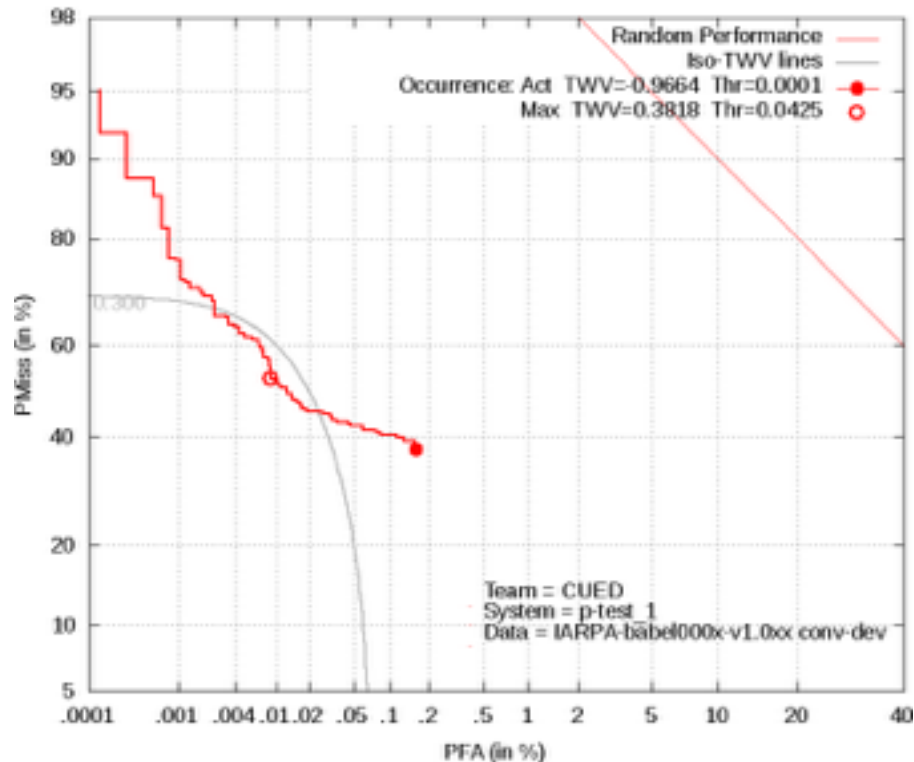
Occurrence scoring, Full Submission, Original orthography, MITLL Force AN. V3



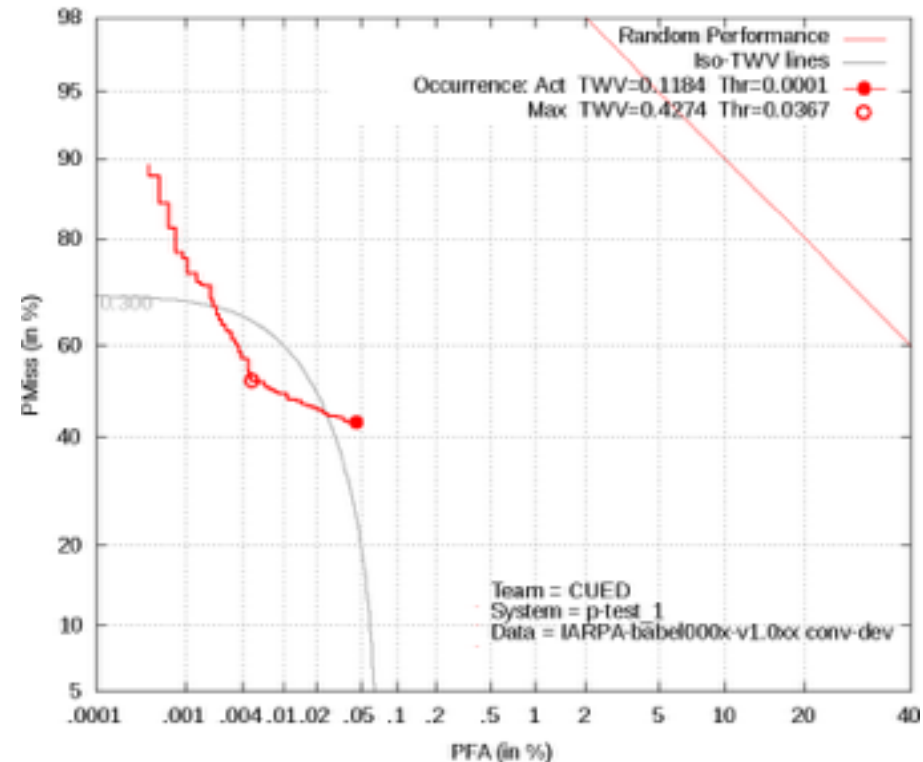
Duty Calls: adding pronunciations

Recall can be increased to 60% by generating pronunciations for the missing words and searching for them in lattices

Occurrence scoring, Full Submission, Original orthography, MITLL Force Ali. V3



Occurrence scoring, Full Submission, Original orthography, MITLL Force Ali. V3



Conclusions

- ASR can be used as the basis for good browsing tools for Oral History
- Using summaries to tune language models is highly effective
- Keyword Spotting from lattices is also effective

Future Work

- Adapt acoustic models for individual speakers
- Links between interviews and between collections
- Add value to the summaries