University of Salford MANCHESTER

VoIPText: Voice Chat for Deaf and Hard of Hearing People

Ben Shirley
Acoustics Research Centre
University of Salford

VoipText

Project commissioned and funded by Ofcom

“Assess if speech to text is sufficiently well developed to make voice chat accessible for deaf and hard of hearing people”

Current Relay Service

- Existing solutions with 3rd party involvement not always appropriate
- Privacy and interruption issues
- ASR solution aims to complement existing services, not replace

VoipText

- Previous work indicates ASR was not sufficiently accurate for telephony with natural speech
- Issues:
  - Natural speech problem
  - Voice specific training

- ETSI ES 202 975 V1.2.1 (2009-10)

Human Factors (HF); Harmonized relay services

Specifies 10% acceptable error rate for ASR – 90% Word Recognition Rate (WRR)

is this a reasonable value?

Speaker Independent Natural Speech Recognition

- Becoming more effective for small vocabulary applications
- ASR from Google and Apple effective for short commands & search terms (with some latency).
- Latency increases with length of input so less effective for natural communication
- Need for *trained* speech recognition

VoipText

Phase 1

•development of speech to text enabled voice chat software based on existing open source voice chat client (GoogleTalk)

Phase 2

•assess performance of two ‘best in class’ ASR engines with reference to ETSI ES 202 975

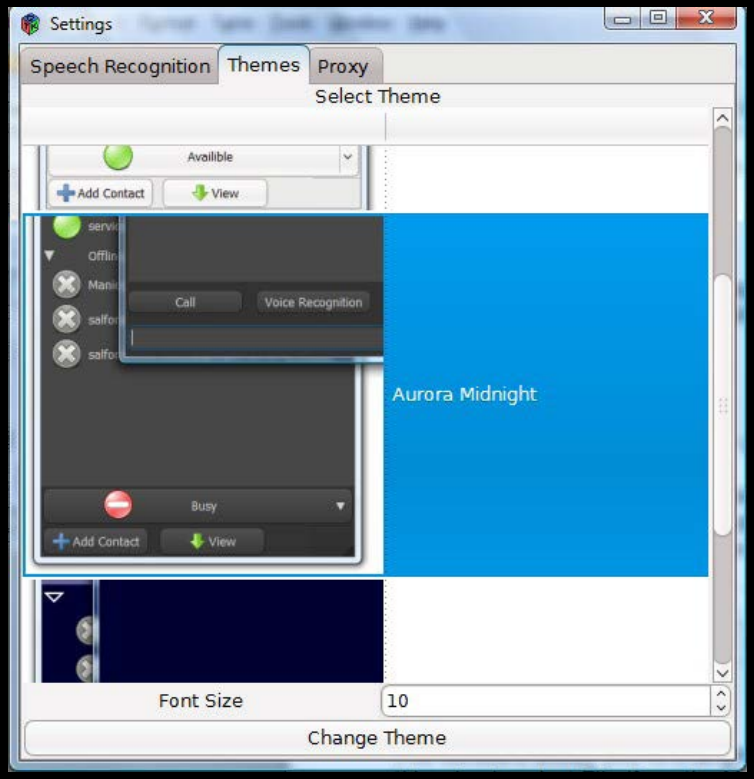
•User assessments and refinement of software

Phase 3

•Extended assessment in users’ homes

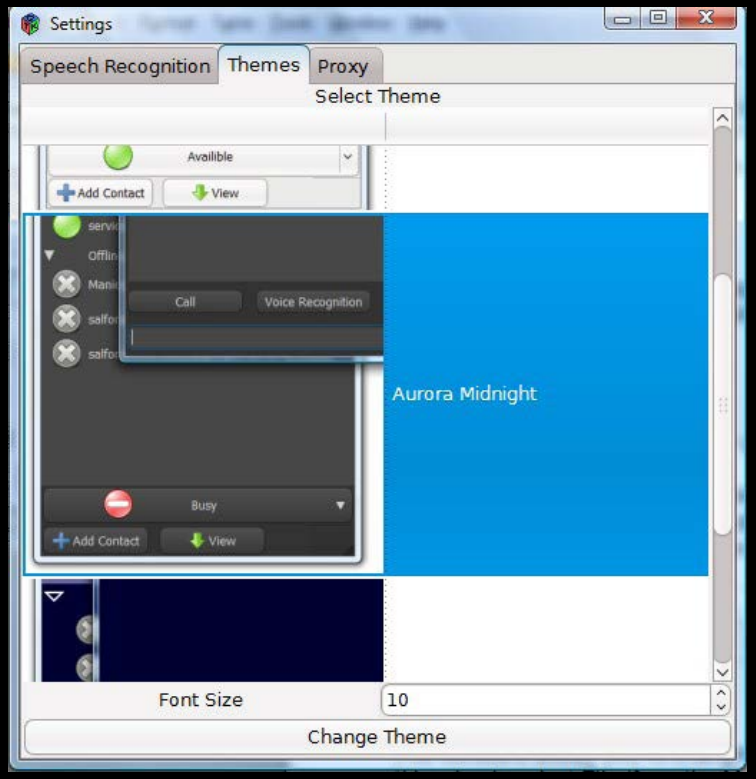
Software Development

- Windows based prototype developed on open source voice chat client
- Developed with input from user groups
- Usability, simplicity, accessibility
- User involvement at several stages in development



Software Development - issues

- Timing of speech to text conversion
  - Contextual recognition issue
- Training requirement – cannot be at HI user's device
  - For HI and non-HI call non-HI person must have installed & trained software



Evaluation of ASR engines

- 6 readers, 3 male, 3 female, various regional accents, *not chosen for their diction*
- Trained each ASR engine (ASR1, ASR2)
- 2 passages of text used
  - Directions
  - Conversation

Example 1: Directions

“*To get to the train station, take your first right, then walk down this road for about 5 minutes. You'll see a large red building on your left and a police station to your right, head right past the police station, and then continue up this road for about 6 minutes. Stop when you come to a coffee shop on your left hand side. You are close to the station now, turn right, then take your second right and continue straight on, you'll be able to see the train station from here, just keep walking straight ahead.*”

Example 2: Conversation

"*Hello Auntie how are you today? Sebastian is now five years old. He starts school this year.*
I hope dinner was nice

We had chicken and rice, it was very nice.
Although the chicken was a little chewy it was still a very nice meal.
I think you can get a cheap flight on August 15th.
It will only cost you 300 pounds which I think is very cheap. Hopefully the weather won't be too bad if you
decide to come it's been raining all winter.
Last week it rained every day although, saying that, it's actually quite nice outside now.
Still, it's not as hot as where you live!
Maybe we should come over there to visit instead.
It would be nice to see Gary and Steve again.
It was nice to talk to you again.
Hopefully we will see you soon.
Give my love to Hannah.
Goodbye!"

Results 1

Example 1: Directions

Subject	*1*	*2*	*3*	*4*	*5*	*6*	Overall
ASR1	96%	79%	79%	80%	81%	80%	82.5%
ASR2	95%	90%	85%	81%	88%	86%	87.5%

Example 2: Conversation

Subject	*1*	*2*	*3*	*4*	*5*	*6*	Overall
ASR1	99%	80%	74%	84%	77%	88%	83.7%
ASR2	92%	89%	86%	74%	89%	75%	84.2%

Overall mean values per subject

Subject	*1*	*2*	*3*	*4*	*5*	*6*	Overall
ASR1	98%	79.7%	76.4%	82.5%	78.9%	85.0%	83.4%
ASR2	93.1%	89.4%	85.4%	76.8%	88.6%	79.3%	85.4%

Results 1

Note subject 1 performance

Example 1: Directions

Subject	1	2	3	4	5	6	Overall
ASR1	96%	79%	79%	80%	81%	80%	82.5%
ASR2	95%	90%	85%	81%	88%	86%	87.5%

Example 2: Conversation

Subject	1		3	4	5	6	Overall
ASR	99%	8%	74%	84%	77%	88%	83.7%
ASR	92%	8%	86%	74%	89%	75%	84.2%

Overall mean values per subject

Subject	1	2	3	4	5	6	Overall
ASR1	98%	79.7%	76.4%	82.5%	78.9%	85.0%	83.4%
ASR2	93.1%	89.4%	85.4%	76.8%	88.6%	79.3%	85.4%

Subject 1 was the most experienced at using ASR.

The users' training may be as important as the ASR engine

Results 1

- Clear benefit to *trained* ASR shown in early pilot tests

- Some evidence that *user experience* may be a major factor.
  - Note results from subject 1 – same time training ASR but more experience in speaking to ASR

Q: Could user training be as important as ASR training?

- Little significant difference between the two leading ASR engines tested
  - only for ‘directions’ example
  - Distinct improvement compared to previous generation ASR

Phase 3 Planning

•Considerable enthusiasm about software from deaf participants

•Both participants and user groups (e.g. RNID) expressed interest in longer term usage study in people's homes

•Phase 3 designed to investigate use of ASR over a period of time with ASR training and user training

•ASR1 chosen for phase 3 (comparable performance with marginal 'best performer' & greater OS compatibility)

Headset choice

•3 commonly used headsets assessed for WRR

•Headset A, WRR = 79.4%
•Headset B, WRR = 93.8%
•Headset C, WRR = 91.8%

•’Best sounding’, most expensive headset was headset A
– different requirements for ASR

•Headset C used for tests
- close to best results and included with the chosen ASR engine so a common choice

Phase 3: User trials

- Carried out over a period of 3 months
- 9 pairs of particpants completed the trial
- Installation on user PC, ASR training and user guidance given on site in participants' homes
- Tech support available via email and phone
- Periodic visits to participants to gain insight into their use of the software and any issues encountered

Phase 3: User trials

Participants' views and experiences collected via:

- Informal documented discussion during home visits
- Focus groups and individual interviews at end of trial
- Questionnaires at end of trial
- Follow up questionnaire

Phase 3: Users' Experiences

- All participants positive in opinion of ASR engine performance

- Easy to follow conversation even where errors occurred

- All participants made substantial use of the software, typically several times a week

- Most popular use would be friends and family in other countries – free calls and more likely to install the software

Phase 3: Users' Experiences 2

Biggest issue was the need to have the trained software on 'the other end' of the call rather than at the HI users' end

•Most usual stated reason why participants would not use the software for 'all' or 'most' calls. "Other people will not want to install and train the software"

•HI users: Considered software to be more useful for people whose voices were unfamiliar but these people would not necessarily have the software installed

•"Would be better if used existing email address or integrated into telephone equipment"

Phase 3: Follow up questionnaire

- 6 of the HI & deaf participants were asked about other applications for ASR in telecommunications including mobile

- 5/6 would use regularly if part of their existing messenger software

- 4/6 would use web based captioned telephony regularly for health information (NHS Direct) and rail timetables

- 3/6 would use web based captioned telephony for banking

And a VoIPText App….?

•Request by participant for mobile app

•Tested – user with GoogleTalk mobile app can receive voice / text from user with VoIPText software

•No additional app required

•although still requiring ASR training at other side of call

•Useful applicationfor access to service providers and public services on mobile

Conclusions

•Overwhelmingly positive response to the ASR engine performance, for these participants ASR *has* reached a useful level of accuracy.

•Need for ASR training means that the implementation used here may not be appropriate

•Positive response to the use of web based ASR captioned telephony for service providers and information services

Mobile

- Smart devices tested at ‘receive’ end of call
- Getting powerful enough to allow effective on-device ASR (e.g. Siri)
- Potential for implementation in social networking apps as well as for service providers and information services
- Other applications???

Thank you

Any Questions?

Ben Shirley
University of Salford
b.g.shirley@salford.ac.uk