



# Transcribing documentaries. Can respeaking be used efficiently?

Lukasz Daniluk<sup>1</sup>, Anna Matamala<sup>2</sup>, Pablo Romero-Fresco<sup>1</sup>  
University of Roehampton<sup>1</sup> & UAB<sup>2</sup>

[lukasz.daniluk@gmail.com](mailto:lukasz.daniluk@gmail.com); [anna.matamala@uab.cat](mailto:anna.matamala@uab.cat);  
[p.romero-fresco@roehampton.ac.uk](mailto:p.romero-fresco@roehampton.ac.uk)

5th Int. Symposium Respeaking, Live Subtitling and Accessibility. Rome, 12.05.15

Funded by FFI2012-31024 & 2014SGR27

# ALST project

- Speech recognition (with/out respeaking)
- Machine translation
- Speech synthesis

In audio description (fiction films)

In voice-over (non-fiction films)

---

# Aim

- Compare three scenarios:
    - Manual transcription
    - Respeaking
    - Automatic transcription + revision
  - Hypothesis:
    - Respeaking could make the transcription of documentaries more efficient
-

# Prior work

- SAVAS, EU-Bridge, Translectures
  - Research presented at previous Respeaking conferences
  - Sperber et al (2013): off-line speech transcription through respeaking via a combination of techniques
  - Bettinson (2013): respeaking in field linguistics (different meaning)
-

# Experimental set-up

- 10 participants (quantitative data from 8, qualitative from 9)
  - Professional transcribers, no previous experience with respeaking
  - 1 video content divided into three 4-minute clips
  - Speech recognition software: DNS 12 Premium
  - ASR transcript generated by EML Transcription server
-

# Experimental set-up

- Background questionnaire (demographics)
  - Training in respeaking (30' theory + 30' practice)
  - Pre-task questionnaire (opinions)
  - Three tasks (randomized): time control and time limit (30' per task)
  - Post-task questionnaire (opinions)
-

# Data obtained

- Quantitative data:
    - time ratio (x minutes transcribing 1 minute of original content)
    - quality of output (NER)
  - Qualitative data:
    - pre-task and post-task opinions on usefulness, speed, accuracy, overall quality
    - post-task assessment of: perceived effort, boredom, confidence in the accuracy of the transcript, and overall quality
-

# Tasks and participants

Number of participants who finished the tasks	ASR	Respeaking	Manual
	3	5	3



# Results: time spent transcribing 1 minute

<b>PARTICIPANTS WHO FINISHED THE TASK</b>	<b>ASR</b>	<b>RESPEAKING</b>	<b>MANUAL</b>
Mean	6'54''	6'26''	5'18''

<b>ALL PARTICIPANTS</b>	<b>ASR</b>	<b>RESPEAKING</b>	<b>MANUAL</b>
Mean	9'36''	8'36''	7'39''

# Results: output quality (NER)

<b>PARTICIPANTS WHO FINISHED THE TASK</b>	<b>ASR</b>	<b>Respeaking</b>	<b>Manual</b>
	98.02	96.88	97.7

<b>ALL PARTICIPANTS</b>	<b>ASR</b>	<b>Respeaking</b>	<b>Manual</b>
	97.535	97.161	97.783

# Summary: objective data (I)

- Manual
    - Fastest method
    - Highest accuracy for all participants, second highest accuracy for those who finished
    - Lower than the 98% threshold for subtitles
  - Respeaking
    - Second fastest method
    - Allowed the highest number of participants to finish
    - Lowest accuracy: no revision
    - Need for specific training
-

# Summary: objective data (II)

- ASR
    - Slowest method
    - High accuracy (built-in revision)
    - Mixed approach
    - More increase in time than in quality
-

# Results: subjective opinions (5-point scale)

Statement	Pre-task	Post-task mean
<b>Manual transcribing</b> is too time consuming	3.4	3.2
<b>Respeaking</b> could be a useful tool to transcribe documentaries	4.5	3.8
Respeaking could speed up the process of transcription	4.5	3.9
Respeaking could increase the accuracy of transcriptions	3.8	2.9
Respeaking could increase the overall quality of transcriptions	3.4	3.1
<b>ASR</b> could be a useful tool to transcribe documentaries.	4.1	2.7
ASR could speed up the process of transcription	4.1	2.1
ASR could increase the accuracy of transcriptions	3.0	2.2
ASR could increase the overall quality of transcriptions.	2.8	2.5

# Results: post-task subjective opinion

	Respeaking	ASR	Manual
<b>Perceived effort</b>	2.89	4.55	3.11
<b>Boredom</b>	2.22	3.89	3.12
<b>Accuracy</b>	2.78	2.89	4.22
<b>Overall quality</b>	3.22	3.00	4.33

# Summary: post-task subjective opinion

- Perceived effort & boredom: respeaking obtains better scores
    - Participants seem ready and willing to try new methods
  - Accuracy and overall quality, manual transcript obtains better scores
    - Habit and familiarity
    - Longer and more tailor-made respeaking training needed
-

# Participants' feed-back

- Impressed with respeaking
  - Need for specific training
  - Combination of techniques (automatic filtering?)
  - Impact on spelling
  - Job satisfaction
    - 88.89 % agreed or strongly agreed that they would enjoy their job more if they used respeaking
    - 11.11 % didn't agree
-



# Conclusions

- First steps towards respeaking for transcription of non-fictional genres
  - Initial hypothesis: potentially more efficient, but need for specific, tailor-made training
  - Better working conditions?
  - Limitations and further research
    - More participants
    - Longer sessions
    - New hands-on tailor-made respeaking method for transcription
    - Automatic system to propose most suitable transcription method
-



# Transcribing documentaries. Can respeaking be used efficiently?

Lukasz Daniluk<sup>1</sup>, Anna Matamala<sup>2</sup>, Pablo Romero-Fresco<sup>1</sup>  
University of Roehampton<sup>1</sup> & UAB<sup>2</sup>

[lukasz.daniluk@gmail.com](mailto:lukasz.daniluk@gmail.com); [anna.matamala@uab.cat](mailto:anna.matamala@uab.cat);  
[p.romero-fresco@roehampton.ac.uk](mailto:p.romero-fresco@roehampton.ac.uk)

5th Int. Symposium Respeaking, Live Subtitling and Accessibility. Rome, 12.05.15

Funded by FFI2012-31024 & 2014SGR27