



Motivation

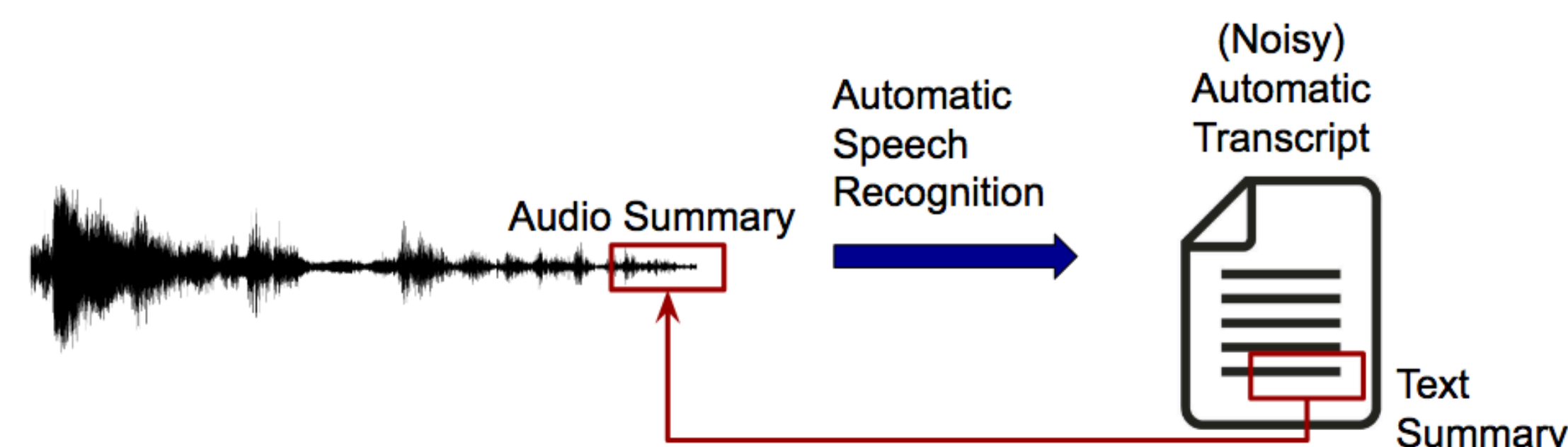
Increase usage of speech-only devices allowing us to use search engines

- How to **present research results** using audio so users can efficiently locate items, determine their relevance, provide feedback, and refine their query if needed?
- How to **structure the conversation** interaction in order to support the user in the information seeking processes with search engines?

Case Studies

Performance and effectiveness of snippets

- **Relevance Judgments:** Can users **identify known-items** from inspecting summaries generated from automatic transcripts?
- **Preference Judgments:** Do users have any **preference for different types of document summaries** (e.g., generated from automatic/manual transcripts)?



D. Spina, J. R. Trippas, L. Cavedon, and M. Sanderson. Extracting Audio Summaries to Support Effective Spoken Document Search. *JASIST*, DOI: 10.1002/asi.23831, 2017.
 J. R. Trippas, D. Spina, M. Sanderson, and L. Cavedon. Towards Understanding the Impact of Length in Web Search Result Summaries over a Speech-only Communication Channel. *In Proc. of SIGIR'15*, 2015.

Form of snippets

- How to better support processing of spoken results lists
 - What is impact of **length** of results summaries?
 - Are shorter spoken summaries as **effective and preferred** as longer more informative summaries



Event Handlers

Gather information about whether:

- The audio was **completely played**
- The audio was **paused**
- The user moved or **skipped** the audio playback to a new position
- **Ranges** of audio have been played or skipped



Challenges

Task Workload

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date
Mental Demand: How mentally demanding was the task?		
Very Low ————— Very High		
Physical Demand: How physically demanding was the task?		
Very Low ————— Very High		
Temporal Demand: How hurried or rushed was the pace of the task?		
Very Low ————— Very High		
Performance: How successful were you in accomplishing what you were asked to do?		
Perfect ————— Failure		
Effort: How hard did you have to work to accomplish your level of performance?		
Very Low ————— Very High		
Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?		
Very Low ————— Very High		

End-to-End Evaluation

Put humans in the loop to interact with speech-only search systems

- **ParIAI Framework**
 - Collect and evaluate conversations between agents and humans via Mechanical Turk
 - Talk to the bots to help train and evaluate them



<http://parl.ai/static/docs/mturk.html>