Speech processing technologies for OSINT and COMINT

Lori Lamel

Vocapia Research & CNRS-LIMSI

Language Technology Industry Summit

Bruseels, October 11, 2017

Vocapia Research & CNRS-LIMSI

- Speech processing research at CNRS-LIMSI since the 1980s
- License of speech processing technology to Vocapia Research in 2000
- Together develop leading-edge multilingual speech processing technologies
- Participation in national and international research projects and technology evaluations
- Many projects related to or funded by defense sector
- Early uptake of research advances (unsupervised learning, neural networks)

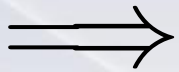
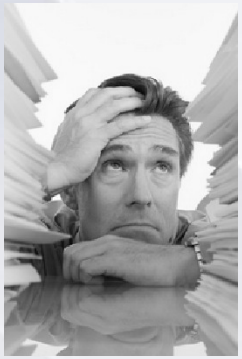
Audio analysis – why?

- Today most information is unstructured (ex. scanned documents, audio, video)
- 300 hours video uploaded every minute (1 hr/s), 54 languages (2017)
- Telephone conversations: over 10 billion calls daily (2013)
- We can store everything but can't really access it

Audio analysis – why?

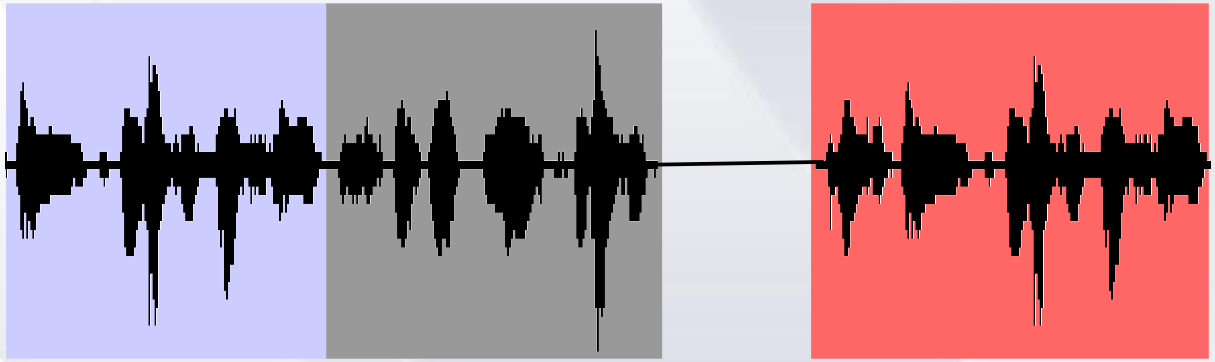
- Today most information is unstructured (ex. scanned documents, audio, video)
- 300 hours video uploaded every minute (1 hr/s), 54 languages (2017)
- Telephone conversations: over 10 billion calls daily (2013)
- We can store everything but can't really access it



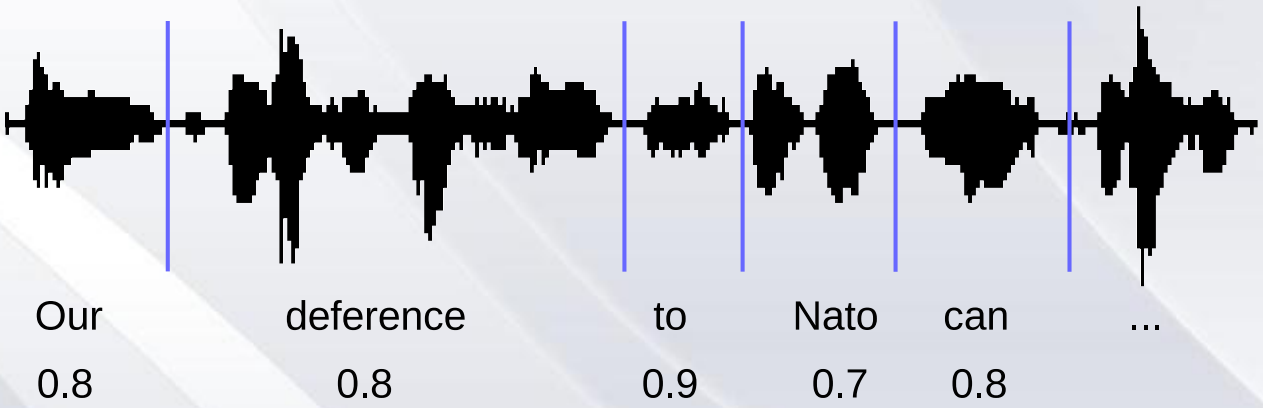
Filter audiovisual data according to a given language, speaker, keywords and/or content ⇒ Improve efficiency and reduce work load of investigators

Audio processing steps



Segmentation	speech	speech	no-speech	speech
LanguageID	English	English		French
Speaker-Diarization	spk1	spk2		spk3
Transcripion				

Speech-to-text transcription



Produce audio transcriptions with

- Time codes
- Confidence measures
- And other metadata: channel, language, speaker, speech turns

Speech-to-text transcription

Why is speech recognition so difficult?

Continuous:	whyisspeechrecognitionsodifficult
Spontaneous:	why'sspeechrecnitionsodifficult
Pronunciation:	wYIzspiCrEkxgnISxnIzsodIfIk∧lt
	wYspiCrEknISNsodIfxk∧l
	wYspiCrEknISNsodIfkL
	....

Speech-to-text transcription

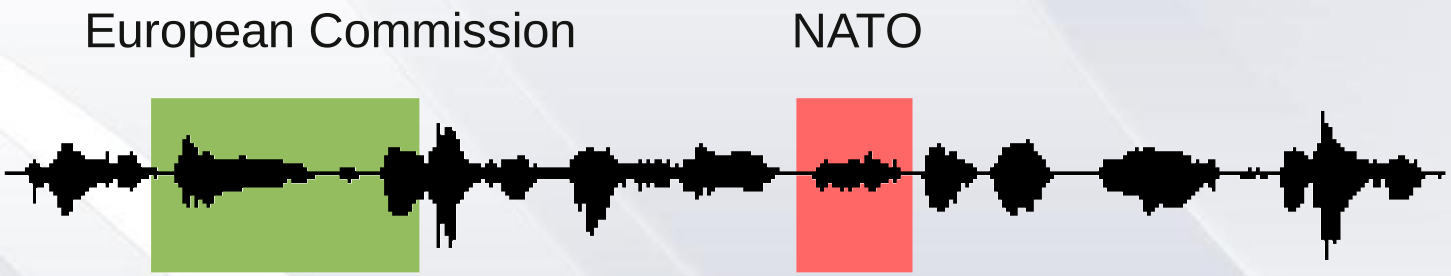
Why is speech recognition so difficult?

Continuous:	whyisspeechrecognitionsodifficult
Spontaneous:	why'sspeechrecnitionsodifficult
Pronunciation:	wYIzspiCrEkxgnISxnIzsodIfIk∧lt
	wYspiCrEknISNsodIfxk∧l
	wYspiCrEknISNsodIfkL

Important variability factors:

Speaker	*Acoustic environment*
physical characteristics (gender, age, ...), accent, emotional state, situation (lecture, conversation, meeting, ...)	background noise (cocktail party, ...) room acoustic, signal capture (microphone, channel, ...)

Spoken term detection



- Locate keywords in an unstructured audio flow
- Different from speech-to-text transcription only target a small set of words
- Trade-off between hits vs false alarms

Core speech technologies

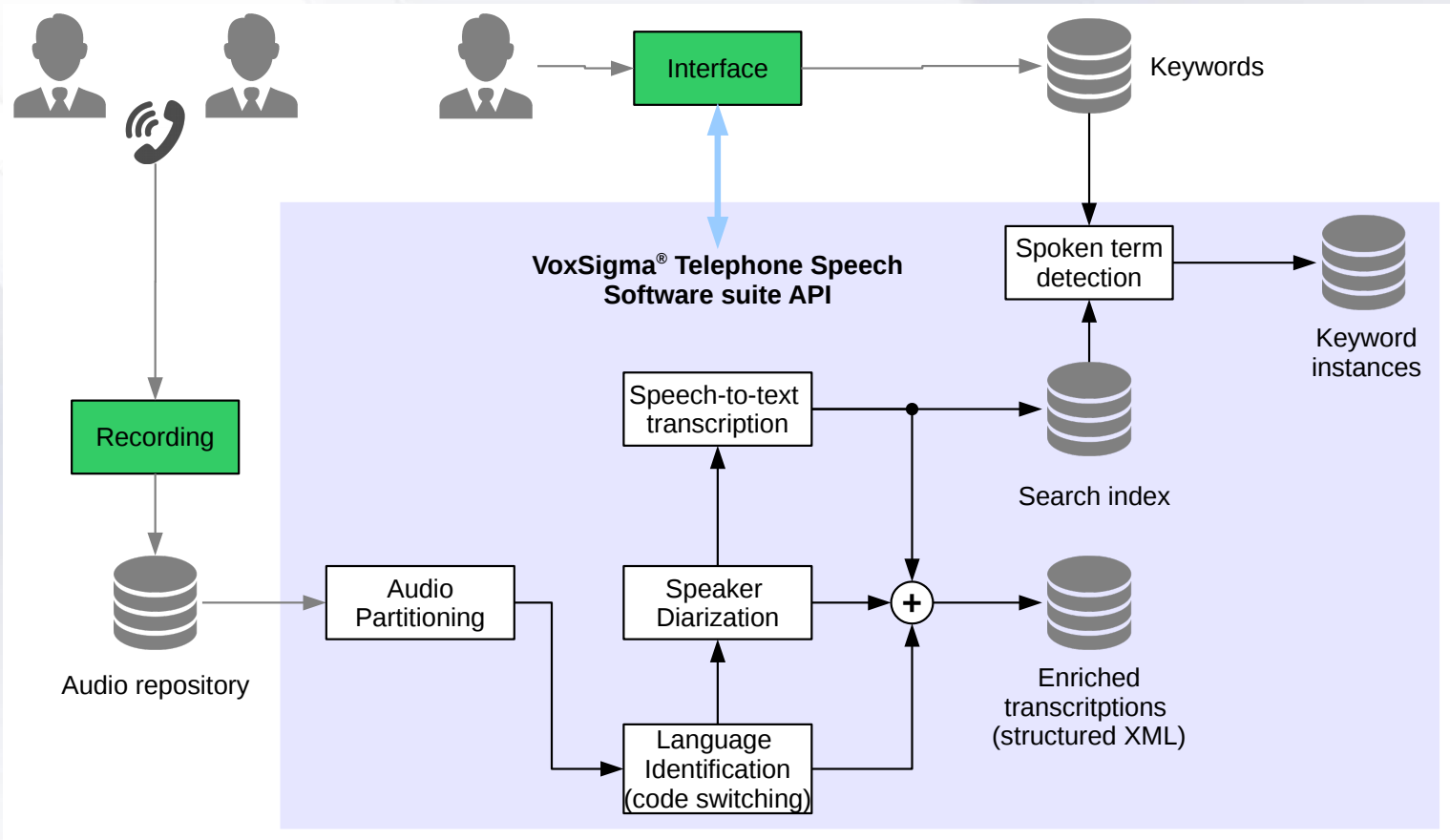
Core speech technologies for **conversational** and **broadcast** speech

↓ COMINT

↓ OSINT

- Speech-to-text transcription and spoken term detection:
  - Conversational speech (11 languages $+20^{+}$):
    English, French, Arabic (Levantine, Algerian, Egyptian), Russian,
    Mandarin Chinese, Spanish, Italian, Dutch, Pashto, Turkish, Vietnamese
  - Broadcast speech (19 languages):
    + German, Romanian, Portuguese, Finnish, Swedish, Latvian,
    Lithuanian, Polish, Greek, Korean
  - Others languages under development
- Language Identification: 48+ languages/dialects

Interface for intelligence analysis of audio



Interface for intelligence analysis of audio

Functional requirements	VoxSigma® features
Detect regions of speech and non-speech, music, continuous/impulsive noises, DTMF, ...	Audio partitioning
Identify (multiple) languages within a recording	Language ID/CS
Convert an audio flow into a structured, searchable and indexable textual data record	Speech-to-text transcription
Detect keywords in audio via textual queries	Spoken term detection

- Defense and commercial applications: call centers (helplines, banking, administration, ...), surveillance
- Human validation instead of exhaustive searching
- Risk of missing an event is reduced

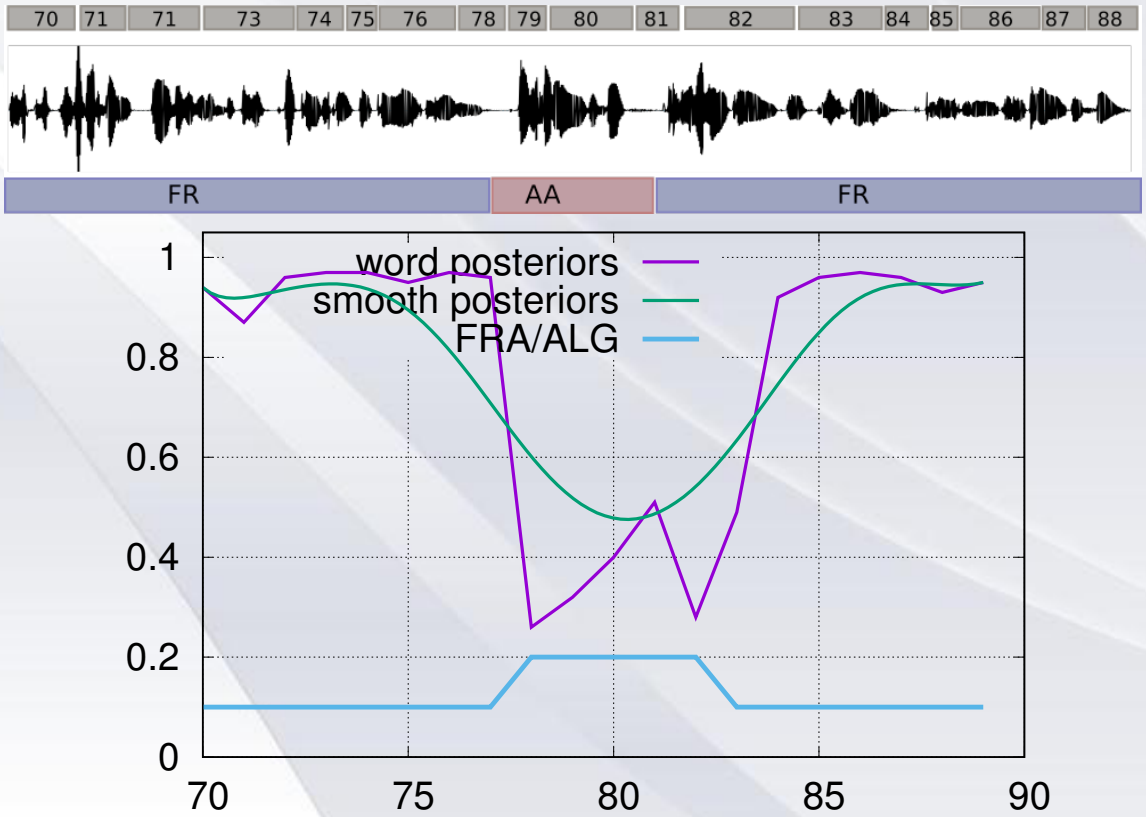
ANR SALSA project - COMINT

Speech And Language technologies for Security Applications
(https://salsa.vocapia.com/)

- **Goal**: to develop set of speech and language processing tools to assist analysts in processing and exploiting audio data for security purposes
- Users: ministry of defense, ministry of the interior
- User needs analysis
  - Audio partitioning for conversational speech (speech activity detection, speaker diarization ("who spoke when"), acoustic event detection)
  - Acoustic model robustness, adaptation to noise, speaker accent
  - Multilingual speech processing, language ID/code-switching
  - Keyword search
- Document selection based on user queries
- Interface to process spoken documents (Scribe) facilitating manual correction

Code switching

- Process of switching from one language to another in a same written or oral conversation (common in some communities)
- Inter-sentential or intra-sentential



H2020 - DANTE - OSINT

Detecting and analyzing terrorist-related online contents and financing activities (http://www.h2020-dante.eu)

- Deliver effective data mining and analytics solutions to detect, collect and analyze multimedia and multi-language terrorist-related contents
- User partners: police (IT, SP, PT, UK)
- Some goals:
  - Accurate and fast detection of suspect terrorists
  - Real-time summarization of terrorist-related content
  - Accurate and fast identification of online terrorist communities
- Audio processing technologies
  - language identification, speech-to-text transcription and audio event detection
  - automate processing of large amounts of data for downstream processing (summarization, translation, named entity detection, ...)

Take home message

- User often don't apprehend properly technological capabilities (too optimistic or too pessimistic, asking too much or too little):
  - transcribing data from native speakers that the listener doesn't understand
  - what acoustic events: all without specifying, including what is said on the TV in the background of a telephone call
- Need close interaction between users and technology developers
- Developers need realistic data targeted by users, and users need to test with real data
- OSINT and COMINT pose different types of challenges
- Technologies have progressed to a point where they serve as an aide to investigators

Thank you for your attention

Questions?

Artist Sven Sachsalber looks for needle in haystack

13 November 2014 | Europe

Share

Italian performance artist Sven Sachsalber is to spend the next two days trying to find a needle in a pile of hay in a Paris museum.

http://www.bbc.com/news/world-europe-30035363

Search shouldn't need to look like this