

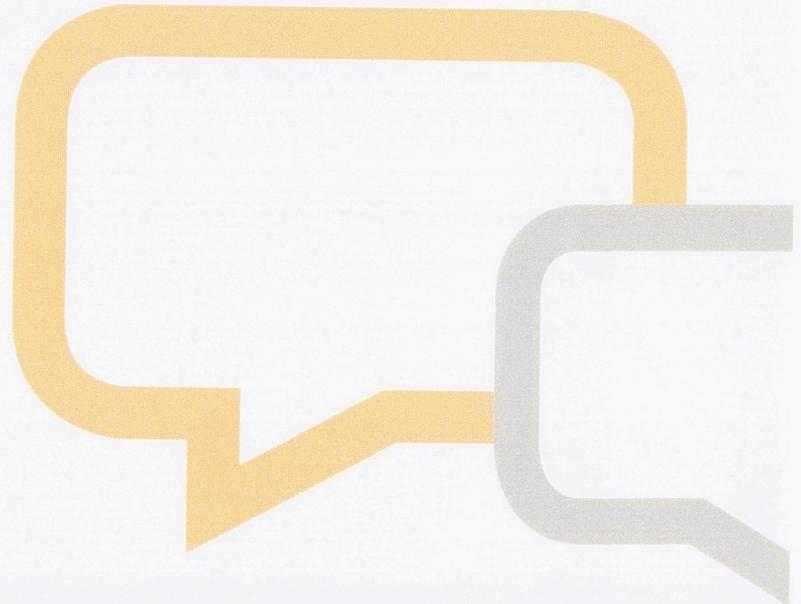
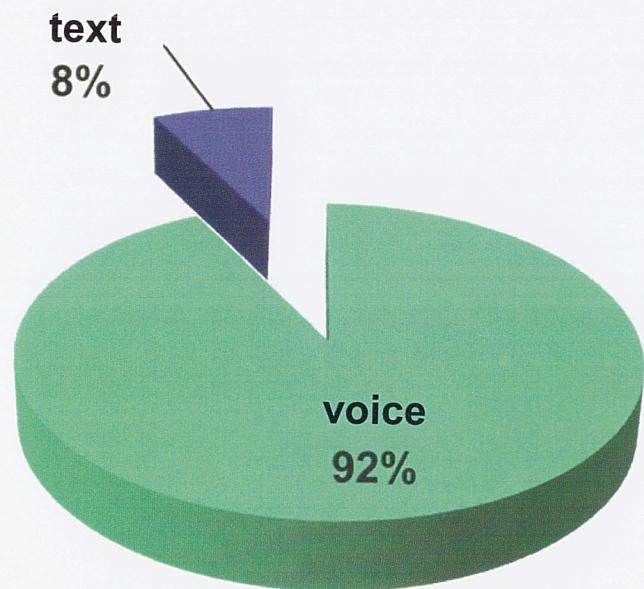


Phone*x*ia

**SPEECH DATA MINING,  
SPEECH ANALYTICS,  
VOICE BIOMETRY**

# INFORMATION IN SPEECH

Average person pronounce about 7400 words a day, but the same person writes only few hundreds of words a day.



Only the text is indexed by search engines, searchable and used in business decisions nowadays. The rest is lost.

# WHAT IS IN SPEECH?

## Speaker

**Language**

**Dialect**, speaker origin

Education

**Gender, age**

**Speaker identity**

## Environment

Where speakers speaks

To whom speakers speaks  
(dialog, reading, public talk)

**Other sounds**  
(music etc.)

## Content

**Keywords**

**Speech transcription**

**Topic**

**When speaker speaks**

## Equipment

Device (phone/mike/...)

Transmit channels  
(landline/cell phone/Skype)

Codecs (gsm/mp3/...)

**Speech quality**

# PHONEXIA

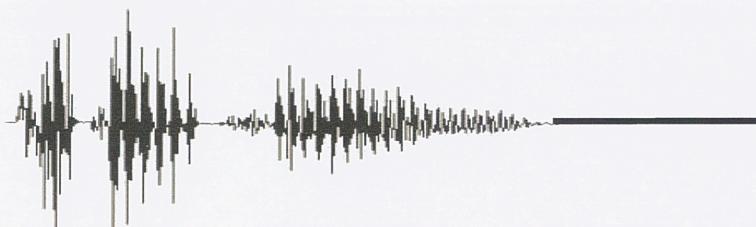
## Goal:

**Help clients to extract automatically maximum of valuable information from spoken speech.**

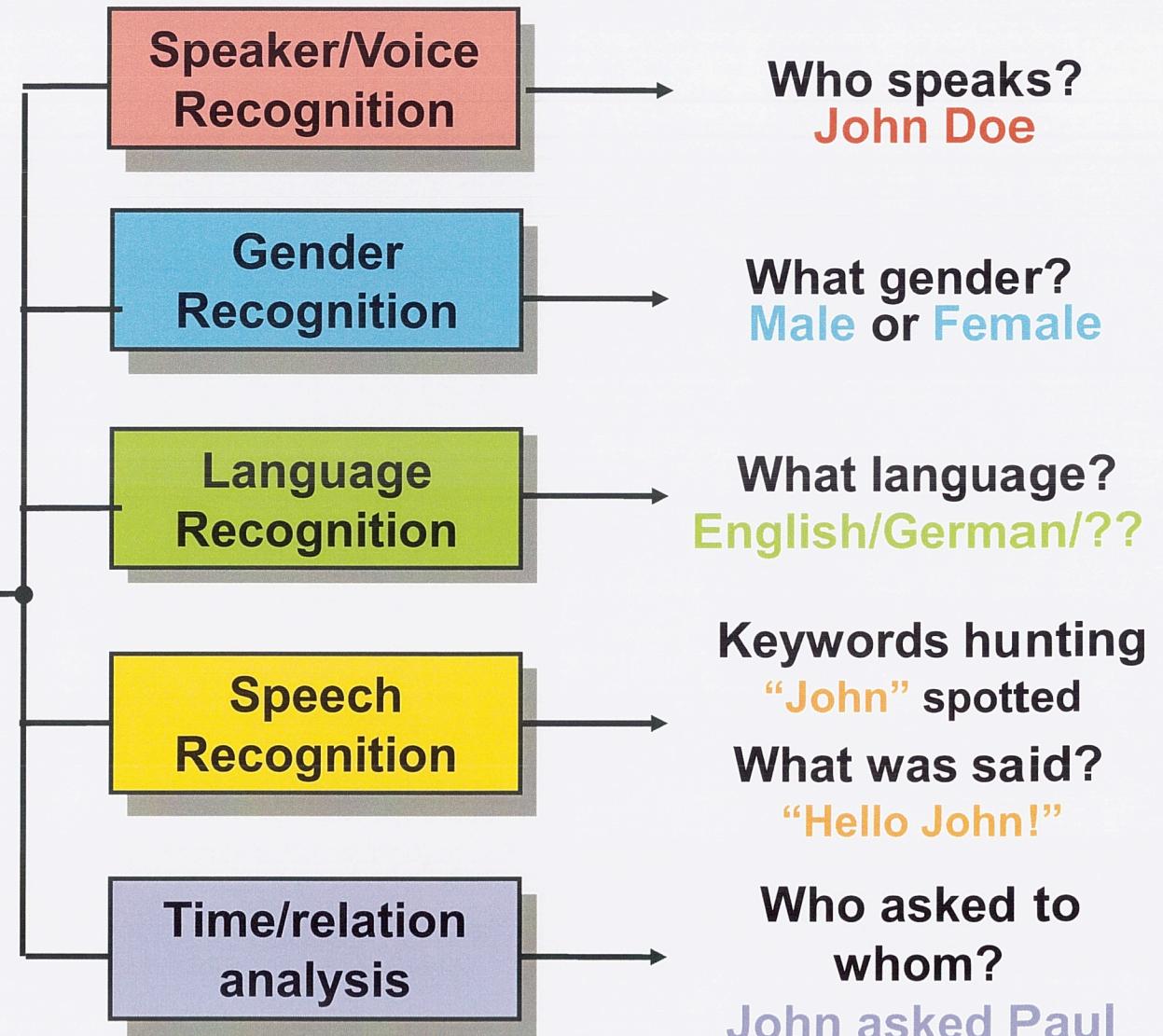
- Based in 2006 as spin-off of Brno University of Technology
- 15 years experience in speech processing, 6 years or work for security/defense sector
- Seat and main office in Brno, Czech Republic, active worldwide
- Tight collaboration with BUT
  - 20 researches moves the technology forward
  - 15 people for development, sales, marketing and technical support
- More than \$1.000.000 invested every year to research and development



# GET THE MOST



audio (speech)



# VALUES AND BENEFITS

## Process automation

- record forwarding, search, authorization of clients

## Higher efficiency

- more records can be processed every day
- spotting of words or phrases, priority queues for analysis by human

## New information

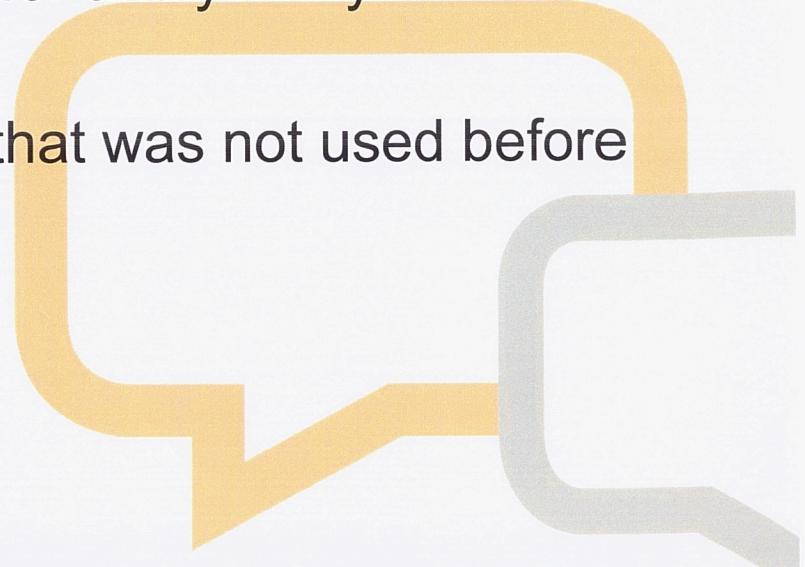
- speech technologies brings new information that was not used before
- analytics of metadata extracted from speech

## Improved quality of services

- rating of agents in contact centers

## Improved security

- voice biometry for authorization of clients on foreground of background
- speaker certifies for trusted calls



# COOPERATION



## We offer:

- **Speech technologies**  
... system design / technology / delivery  
/ technical support / maintenance
- **Consultations**
- **Custom development and research**

# CLIENTS AND PARTNERS

## Customers and partners

- Security and defense agencies
- Call centers
- IT integrators
- HW suppliers
- TV and radio stations (audio archive indexing)
- Universities
- Customers on several continents:  
USA, Russia, Germany, France  
Switzerland, Czech Republic,  
Poland, Slovakia, Spain, Romania,  
Israel, India, Mexico ...



Ministerstvo obrany  
České republiky

OptimSys



SIEMENS



TOVEK



MEDAV



INEXIA

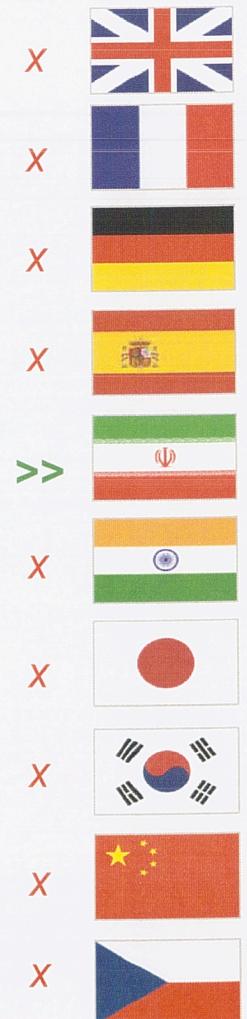


# LANGUAGE RECOGNITION

- Automatic recognition of the language spoken.
- 50 languages + user can add new by himself!
- Can be used also as dialect recognition
- Acoustic channel independent
- iVector based technology, discriminative training, < 1kB voiceprints

## Usage:

- Crime is caused by some minorities very often
- Call record forwarding  
(to operator / other technologies / archive ...)
- Analysis of the audio archive
- Monitoring of some call services or media sources.



# GENDER RECOGNITION

- Automatic recognition of gender
- Highly accurate and robust technology, very easy deployment
- Acoustic channel and language independent
- Extremely fast

## Usage:

- Narrow the searched space to half or quarter (dialogs)

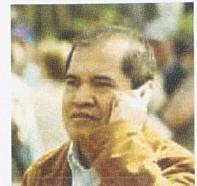


# SPEAKER RECOGNITION

- Several scenarios: speaker verification, speaker search, speaker spotting, link/pattern analysis
- Acoustic channel and language independent technology, noise and channel distortion robust technology
- Based on iVectors
- Voiceprint extraction and scoring
- Compact voiceprint representation (about 600 bytes)
- Millions of comparisons in fraction of seconds
- Diarization (speaker segmentation)
- User based system training, user based calibration



>>



X



X

# LINK ANALYSIS

Explore relations among people  
if they use  
several phones

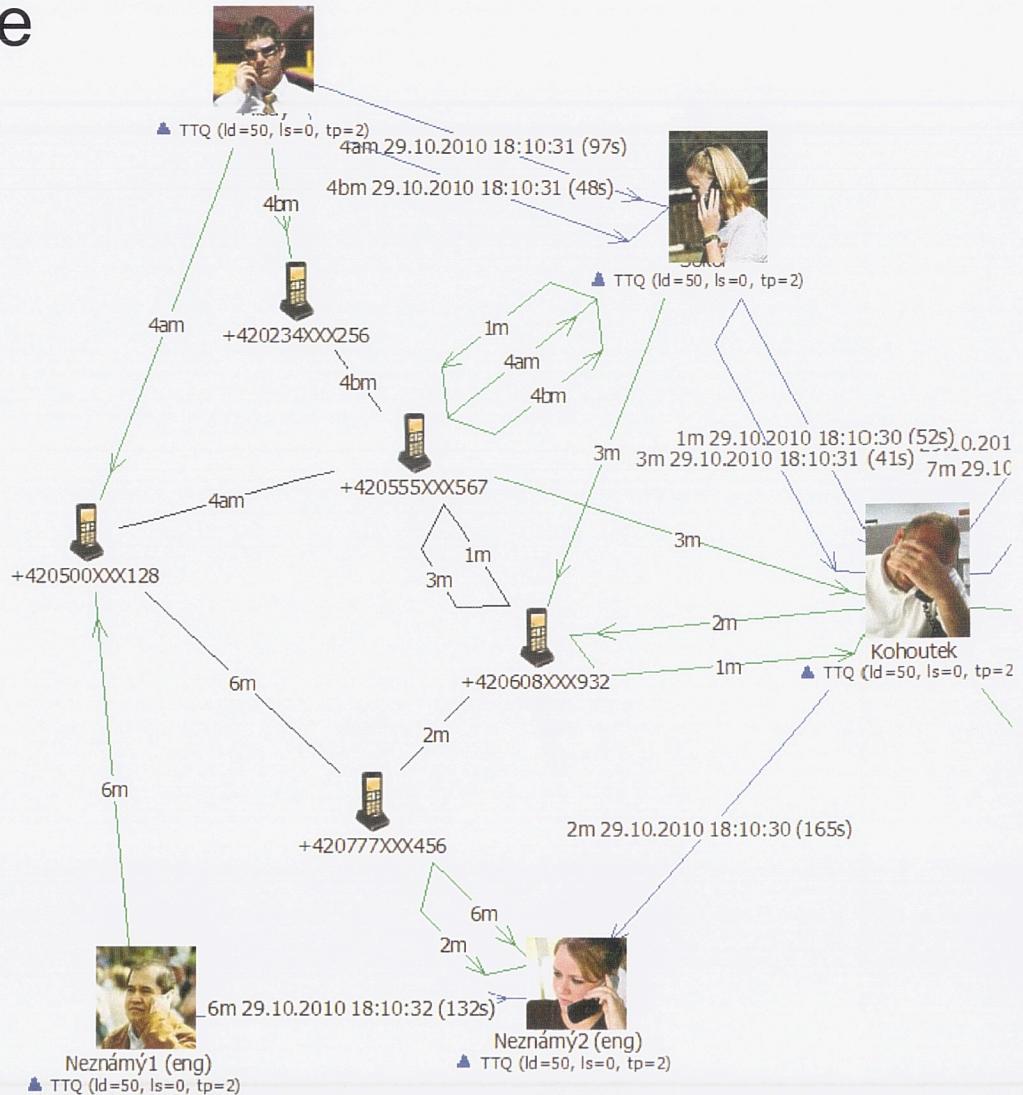


Link analysis tools

## Usage:

- Search for patterns
- Time analysis
- Frequency of calls proceeded by specific people

(I2 Analyst's Notebook)



# ANDROID SPEAKER RECOGNITION



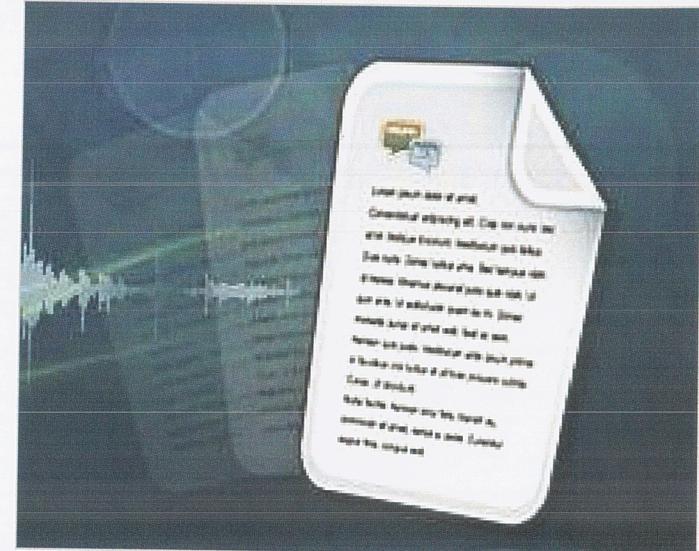
- Easy identification of suspects in streets
- Verification of collaborators based on voice biometry
- Runs fully on cellphone
- Voiceprints can be shared with large landline/GSM/VoIP ... monitoring systems
- Possibility to create voiceprint directly on the device
- Several thousand voiceprints fits easily

# SPEECH TRANSCRIPTION

- Information can be searched quickly anytime
- Transcript can be processed by tools for content analysis of text
- Reading is faster than listening
- Alternative transcription hypothesis  
-> almost 100% of spoken words can be found
- Languages: EN (native/non-native), CZ, RU

## Online example:

- SuperLectures.com – search in lectures



# SPEECH TRANSCRIPTION

Speech



Source

What if they did something similar to what they do  
AT the Vietnam memorial. The Vietnam Memorial is  
staffed AND run by a lot of THE volunteers who are  
responsible for...

Transcript

what if they did something similar to what they do  
WITH the vietnam memorial the vietnam memorial is  
staffed IN run by a lot of \*\*\* volunteers who are  
responsible for...

Message in newsletters

# INTEGRATION WITH TEXT BASED TOOLS - TOVEK SERVER

- Processing of large volumes of speech transcriptions
- Indexing and search, complex queries
- Search in alternative recognition hypothesis
- Categorization of records
- Delivery of content based on user defined profiles (sets of queries)
- Metadata (comments, knowledge ...)



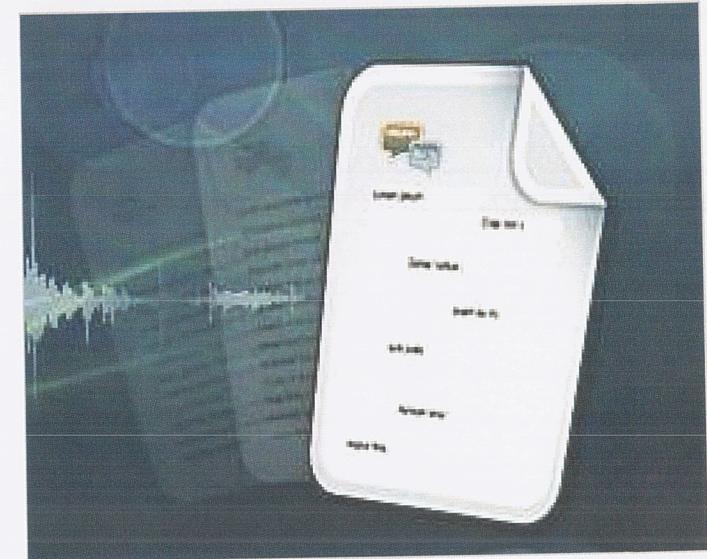
TOVEK

# KEYWORD SPOTTING

The input is a list of keywords, the output are occurrences with scores

## Speech transcription based KWS

- speech transcription including alternative hypothesis and word confidences, comparison of confidences with threshold
- Very accurate, but slower and expensive to develop
- Can be indexed easily

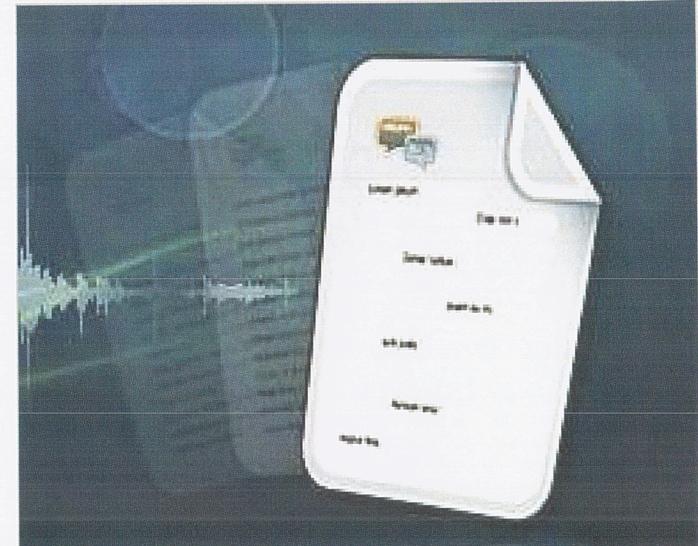


## Acoustic KWS

- can be developed quickly for any language with low cost, less accurate

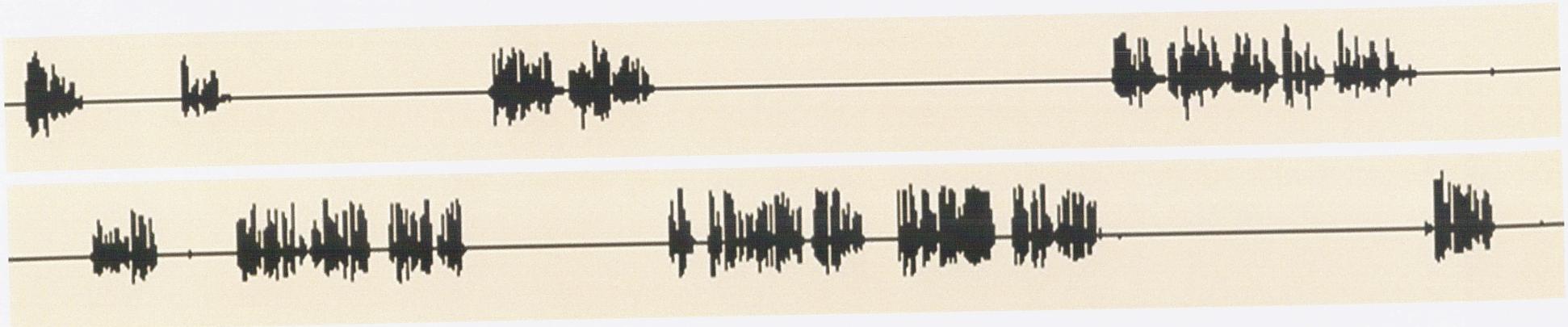
# ACOUSTIC KEYWORD SPOTTING

- Fast analysis of the content for pre-selection of records
  - Often used in quality control in contact centers
  - Keywords entered as a text are searched in the audio
  - Languages: EN, RU, GE, PL, CZ, SK, HU, Levantine AR, Venezuelan SP
  - New languages are added every year
  - New language can be added as a service in 1 to 2 months  
(10 to 20 hours of annotated speech is necessary)



# TIME ANALYSIS OF DIALOGS

- Can help to preselect highly informative records for analysis
- Many statistics: speaker activity, number of speaker turns, speech speed, information flow, cross talks, reaction times
- Statistics over source person or time



# INTEGRATION (HIGH LEVEL)

- GUI desktop version
  - evaluation, simple use cases
  - specialized applications
- Command line
  - easy integration in scripting languages (Perl, Python, Bash, Tcl/Tk ...)
  - detailed evaluations
- Intelligence Platform
  - Integration of all speech technologies, processing schema in XML, output metadata in XML

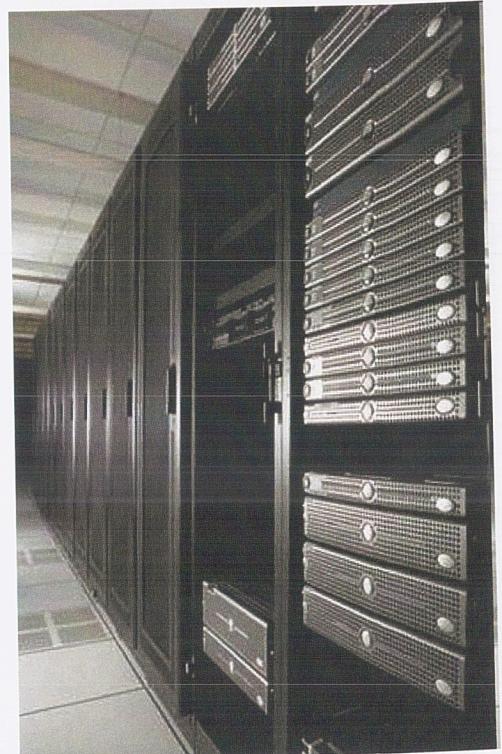
# SPEECH INTELLIGENCE PLATFORM

- **Situation**

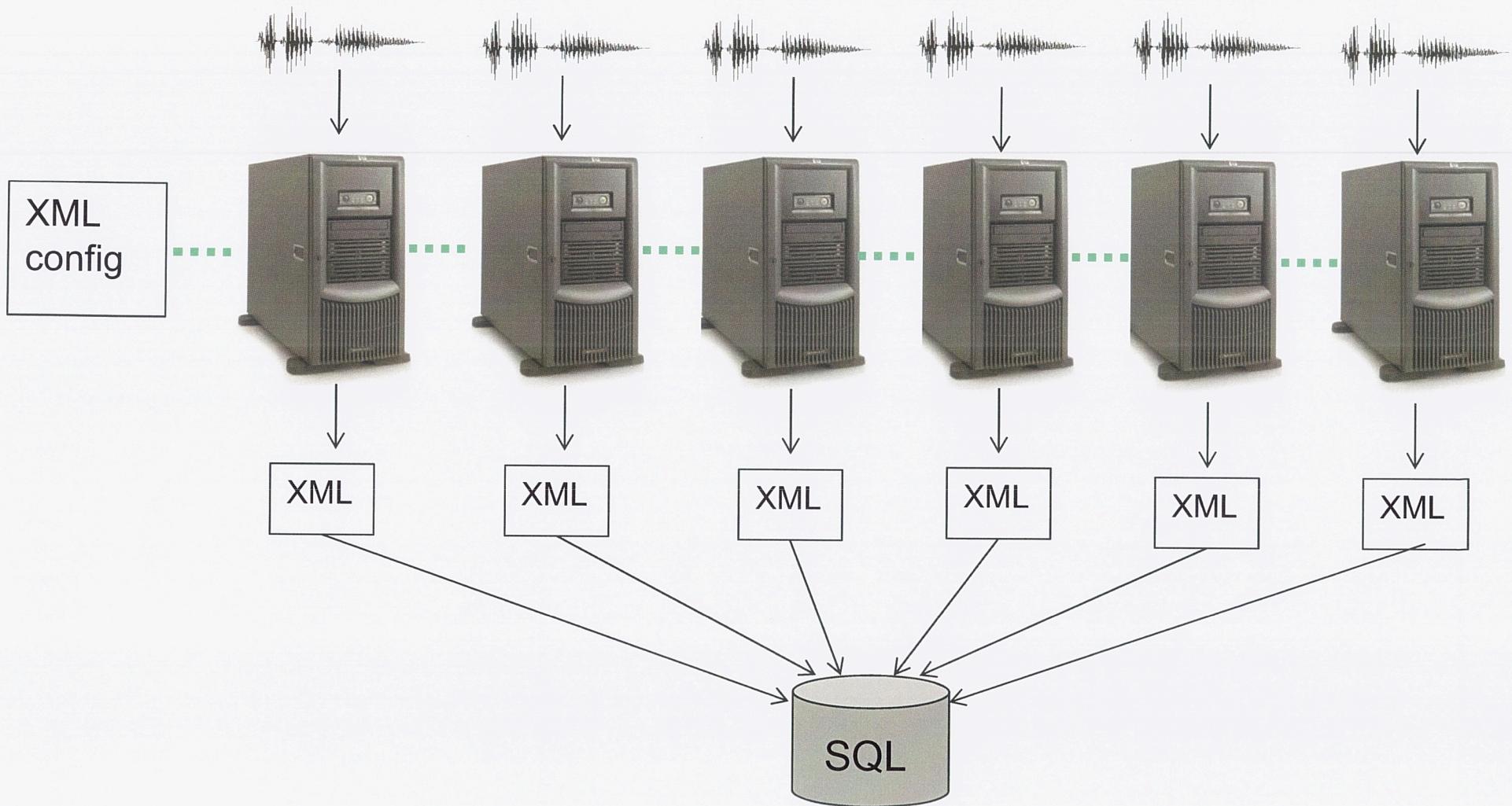
- Data mining systems for intelligence
    - many technologies involved,
    - data come from several sources
    - processing schemes change rapidly
    - exact procedures are secret know-how.

- **Solution = Speech Intelligence Platform**

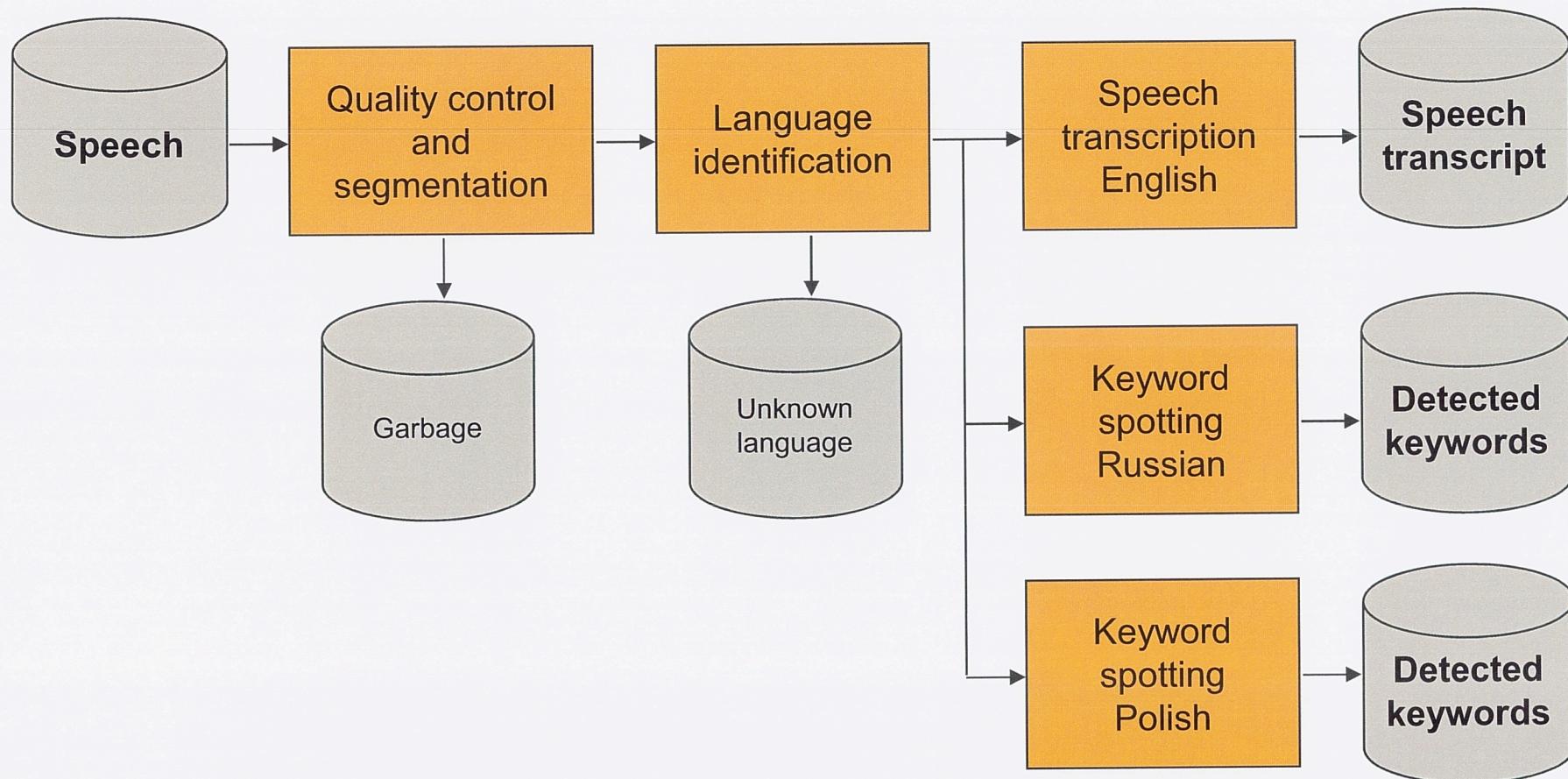
- Design and deployment of the speech processing system alone, in very short time and without deep knowledge about speech technologies.



# SPEECH INTELLIGENCE PLATFORM



# MULTILINGUAL SPEECH TRANSCRIPTION AND KEYWORD SPOTTING SYSTEM



# INTEGRATION (LOW LEVEL)

- Software + SDK (API, documentation, examples)
  - for direct integration to client's IT system
  - C++ API, Java API (TomCat, JBoss, desktop apps), C# API
  - platforms Windows/Linux 64/32 bits, Android
  - easy porting to any POSIX platform that supports GCC
- W3C network protocols (MRCP v2/SIP/RTP) for integration with common IVR platforms (Avaya, Cisco, OptimTalk, FreeSwitch, Asterix ...) for Speaker ID

# SPEECH ANALYTIC SERVER

CheckRecordPage +

Speech Analytics Server

Cards | Temporary records | Statistics | Analytics | Settings | About | Logout

Record input

o Saved file: david\_1.wav

Select file: Procházet... Upload 00:00 00:00

Uploaded record

File	david_1.wav	Name		Insert date	05.06.2012	Language	English	<a href="#">Update this record</a>
Owner	Milan	Length [s]		Speech length [s]	17	Gender	Male	

Compare result

Showing 1 to 10 of 1015 << < 1 2 3 4 5 6 7 8 9 10 >>

	Audio	Probability (%)	First name	Last name
1	[audio player] 00:00 1:00:00	99	Quentin	Straley
2	[audio player] 00:00 1:00:00	99	Nitin	A
3	[audio player] 00:00 1:00:00	99	David	Bowie
4	[audio player] 00:00 1:00:00	98	Malcolm	Tillotson
5	[audio player] 00:00 1:00:00	73	Buster	Herbst
6	[audio player] 00:00 1:00:00	72	Silas	Coelho
7	[audio player] 00:00 1:00:00	69	Miles	Polen
8	[audio player] 00:00 1:00:00	65	Miquel	Siegfried
9	[audio player] 00:00 1:00:00	49	Neville	Camire
10	[audio player] 00:00 1:00:00	44	Alphonso	Augustus

Add the record to the selected speaker | Create new speaker with the record

Speech Analytics Server

Cards | Temporary records | Statistics | Analytics | Settings | About | Logout

Min. reaction time (of channels)

Count of the records

Legend: Zoom in: Click and Drag  
Zoom (back) and Double-Click  
Shift the chart: Shift + Click and Drag  
Rolling average: Text box in the lower left-hand corner of the chart  
Select records for Analytics: Ctrl + Click on the Point  
Select range of records for Analytics: Ctrl + Click and Drag

Max. reaction time (of channels)

Count of the records

Legend: Zoom in: Click and Drag  
Zoom (back) and Double-Click  
Shift the chart: Shift + Click and Drag  
Rolling average: Text box in the lower left-hand corner of the chart  
Select records for Analytics: Ctrl + Click on the Point  
Select range of records for Analytics: Ctrl + Click and Drag

Average reaction time (of channels)

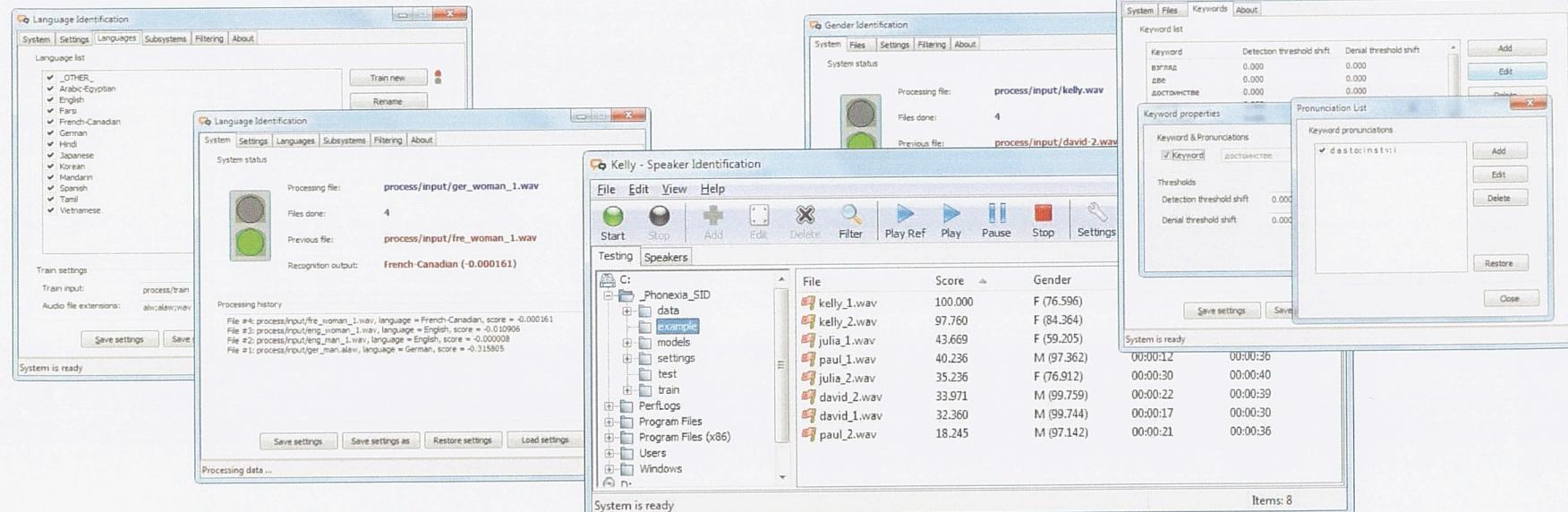
Count of the records

Legend: Zoom in: Click and Drag  
Zoom (back) and Double-Click  
Shift the chart: Shift + Click and Drag  
Rolling average: Text box in the lower left-hand corner of the chart  
Select records for Analytics: Ctrl + Click on the Point  
Select range of records for Analytics: Ctrl + Click and Drag

# EVALUATION

[www.phonexia.com/download](http://www.phonexia.com/download)

- Software for Windows/Linux in GUI and cmd-line versions
- Free evaluation license file is sent to email



# HOW CAN WE HELP?

**Phonexia is  
a technology company**

We can offer:

- Consultations
- Speech technologies and solutions
- Custom development.
- Research

The screenshot shows a website interface for Phonexia. On the left is a sidebar with links: Company, Technologies, Products & Services, Phonexia, Speech@IT, References, People, and Contacts. Below this is a login form with fields for User and Password, and a Logon button. The main content area has a header "Download Phonexia Demo Applications". It lists four demo applications with descriptions and screenshots:

- Phonexia Language Identification from Spoken Speech Demo: Describes identifying language spoken in calls.
- Phonexia Speaker Identification System Demo: Describes identifying speakers in speech records.
- Phonexia Acoustic Keyword Spotting Demo: Describes spotting keywords in speech records.
- Phonexia Gender Identification Demo: Describes identifying gender in calls.

At the bottom right of the content area, it says "© Phonexia 2008-2010 | last updated: 15.04.2010 | info@phonexia.com".

**We help our clients  
to get the most from  
speech records**

# Q & A



**Phonexia s.r.o.**

+420 511 205 265

+420 733 532 890

[info@phonexia.com](mailto:info@phonexia.com)

[www.phonexia.com/download](http://www.phonexia.com/download)