

# Complex Identification Decision Based on Several Independent Speaker Recognition Methods

Ilya Oparin  
Speech Technology Center

## Global provider of voice biometric solutions

**Company name:** Speech Technology Center, Ltd

**Core expertise:**

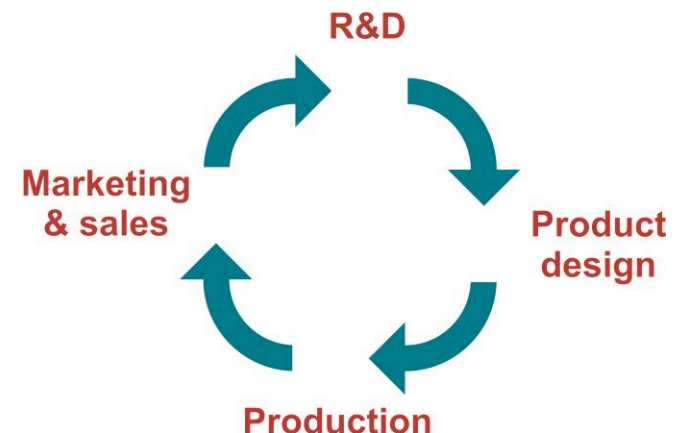
Voice identification and verification	Professional audio recording
Audio forensics	Noise cancelation

**Location:**

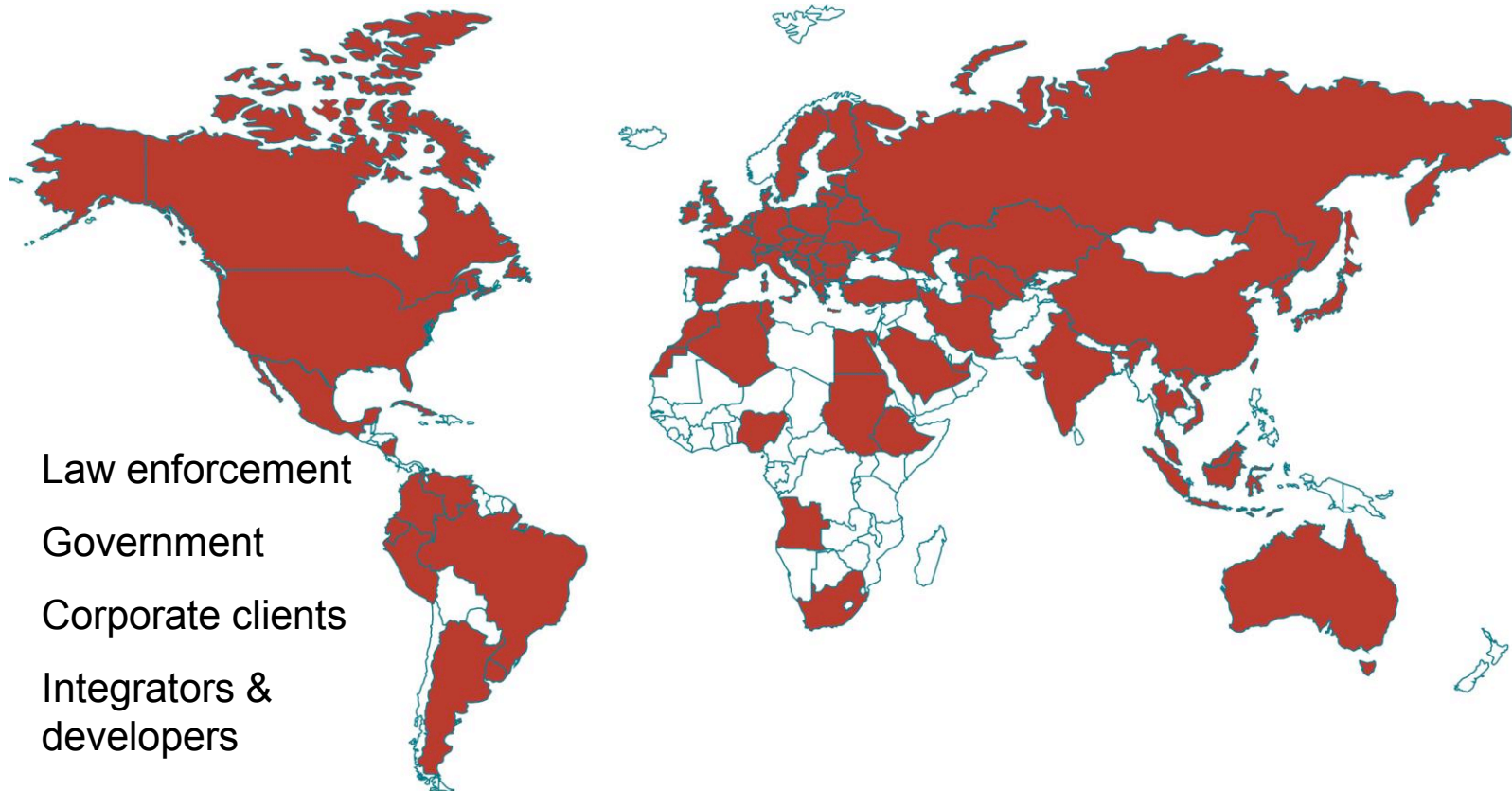
Russia  
Germany  
Mexico  
USA (office in 2009)

**The year of foundation:** 1990

**Staff:** 250 including 25 world-class PhD



# Global Customer Base in More than 60 Countries

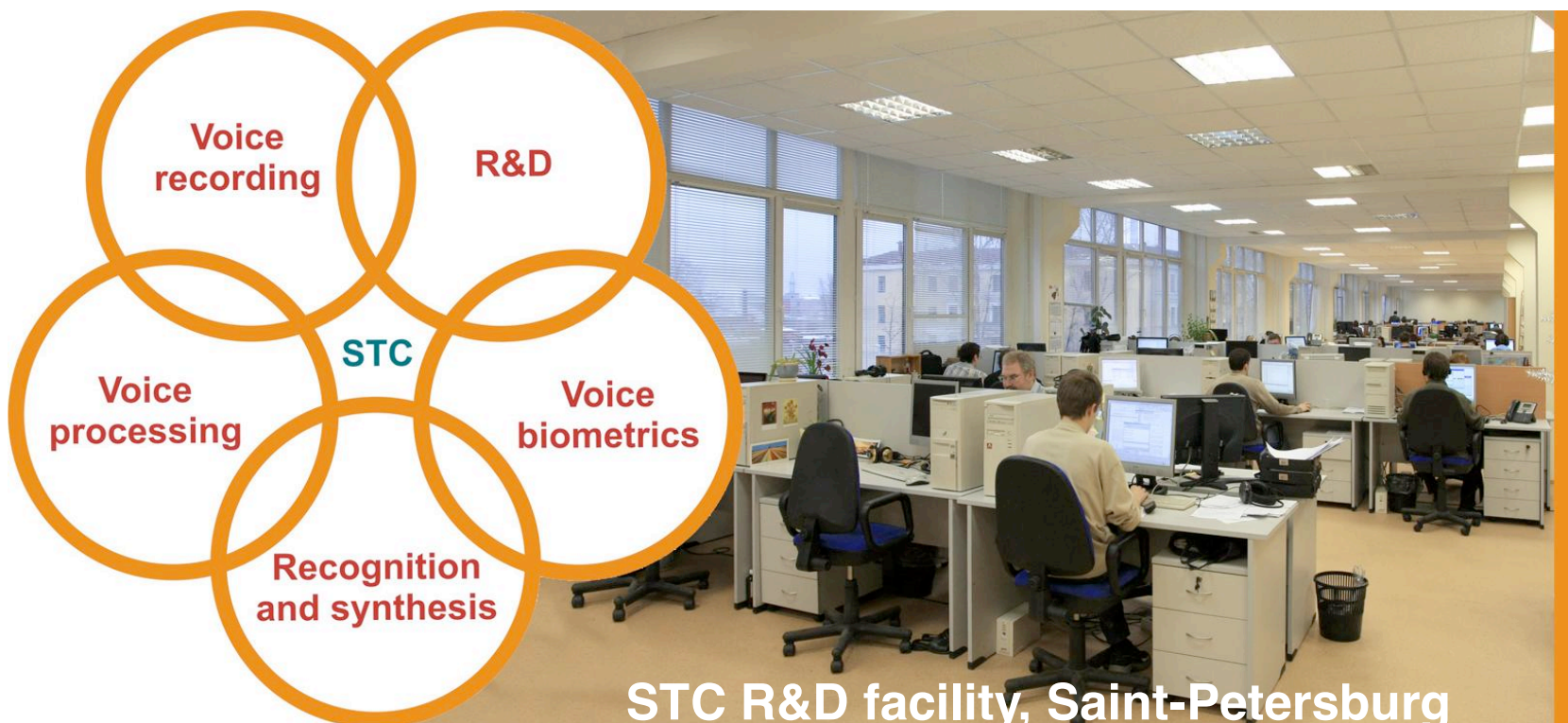


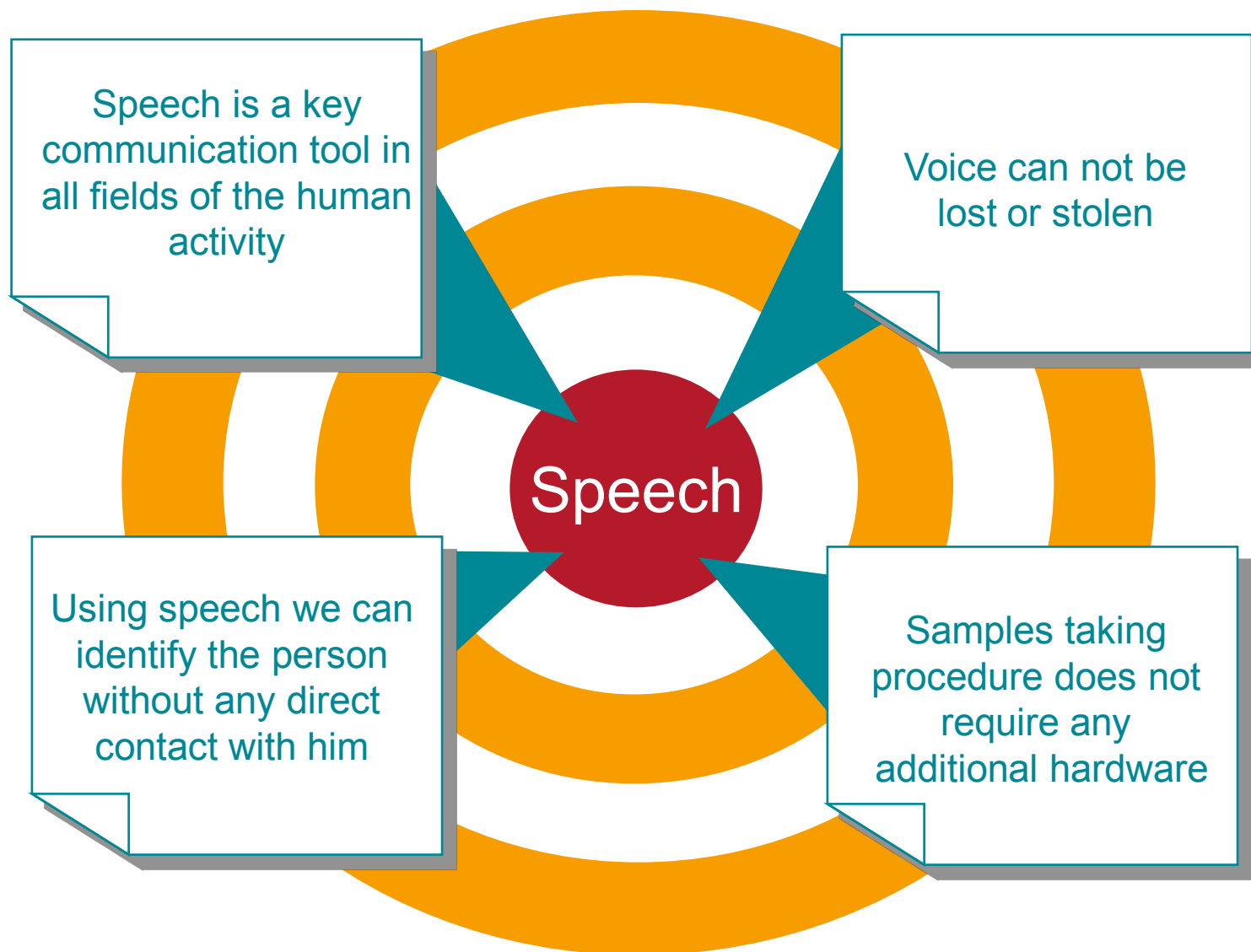
- ➔ Law enforcement
- ➔ Government
- ➔ Corporate clients
- ➔ Integrators & developers



### Ambitious and experienced team:

- One of the leading R&D teams (voice sector) in the world: over 100 technical specialists, scientists and software developers (including 25 PhDs), 5 certified audio forensic experts.
- Strong management and sales teams







## Global leader in audio forensics Over 15 years of experience

- ➔ Forensic speaker identification.
- ➔ Authenticity analysis of analog or digital audio recordings.
- ➔ Audio equipment for forensic examination and identification.
- ➔ Speech enhancement and audio restoration.
- ➔ Text transcription of low quality recordings.

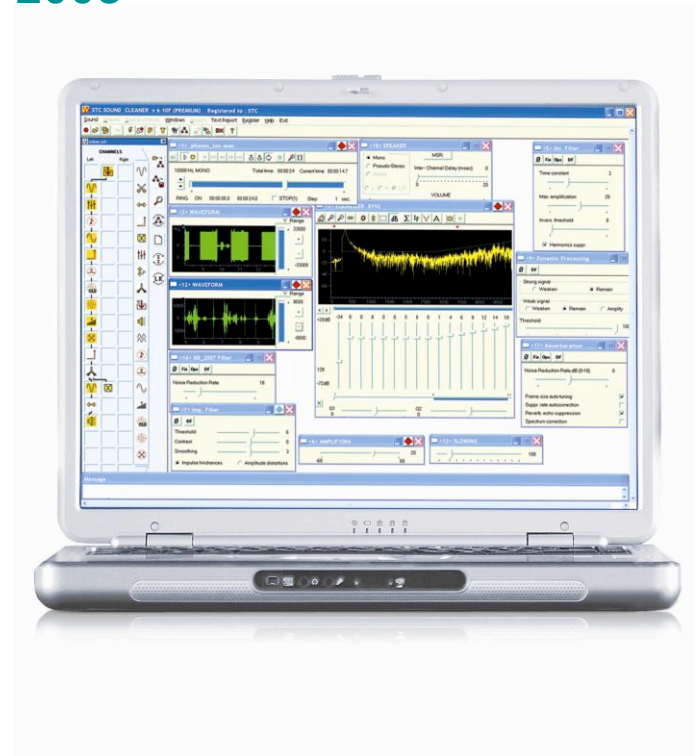


- ➔ **Automatic algorithms for real-time noise suppression and speech enhancement.**



Sound Cleaner Premium – the first and the second prize in audio enhancement contest by AES (Audio Engineering Society), Denver, 2008

- ➔ Efficient suppression of all types of noises and distortions
- ➔ Adaptive algorithms of filtering
- ➔ Filters can be combined to process the record simultaneously



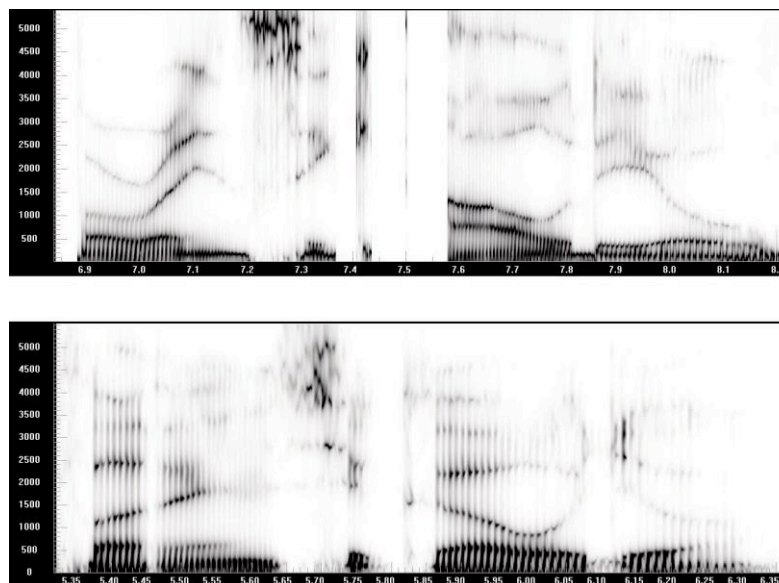
## State-of-the-art voice-ID systems face four basic challenges:

- ➡ Ensuring robustness to noise (real life audio)
- ➡ Ensuring robust performance across different sound recording channels and levels of speaker stress
- ➡ Effective processing of large-scale (nation-wide) databases
- ➡ Language and context independent identification



## Spectral-formant method

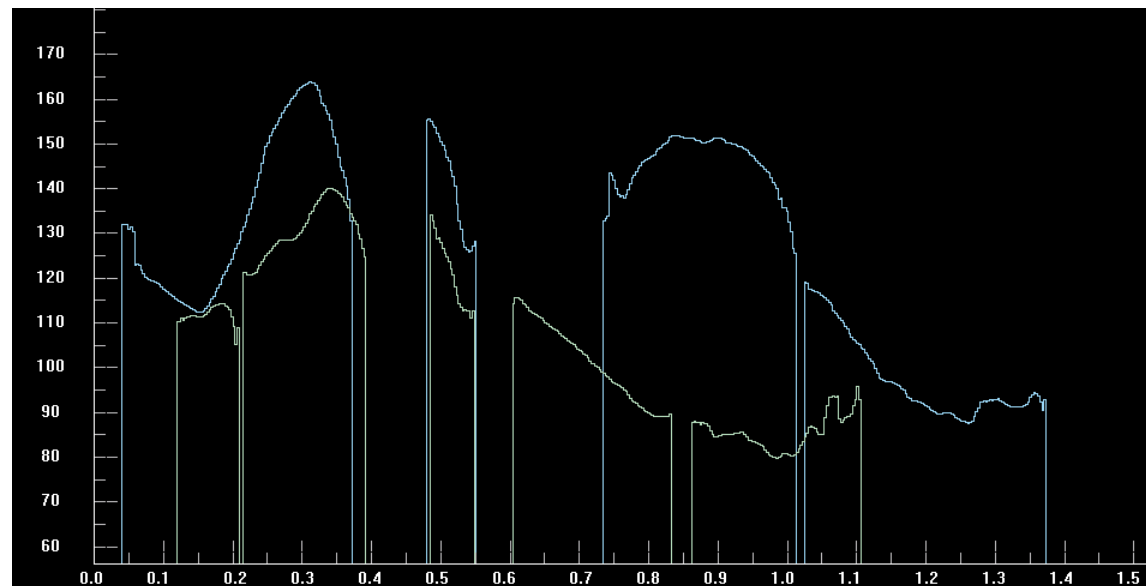
- ➔ Spectral-formant method (SFM) is based on the unique shape of each person's vocal tract which is reflected in the visible speech of different people.



An example of formant representation of the phrase “Forensic audio” pronounced by two different persons is shown in the picture (The horizontal axis is time in seconds. The vertical axis is frequency in Hertz. Energy level is depicted by the darkness of the trace).

## Pitch statistics method

- ➔ Pitch statistics method (PSM) engages 16 different pitch parameters, including average pitch value, maximum, minimum, median, percent of areas with rising pitch, pitch logarithm variation, pitch logarithm asymmetry, pitch logarithm excess and 8 parameters more.



An example of automated pitch extraction in the phrase “Forensic audio” pronounced by two different persons is shown in the picture

## GMM/SVM method

- ➡ In the GMM/SVM approach Gaussian mixtures are used to approximate statistical distributions of MFCC (Mel frequency cepstral coefficients) parameters extracted from speech of different speakers.
- ➡ Support Vector Machines are a robust classifier in multi-dimensional space.

Method	Dependence on speech signal characteristics		
	Signal duration	Signal quality	Emotional state
Spectral-Formant	+	++	+++
Pitch Statistics	++	+++	+
GMM/SVM	++	+	++
Fusion (STC)	++	+++	+++

### **Ability to work with signals from various communication channels**

Both microphone and telephone (landline, GSM)

### **Robust to noise**

Low-quality signal processing (SNR down to 10 dB)

### **Processing of short speech signals**

Speaker identification by a few seconds of speech

## Database

NIST SRE 2004

## Spectral-Formant method

EER=13%

## Pitch statistics

EER=15.9%

## GMM/SVM

EER=7.5%

## Fusion

EER=4.7%

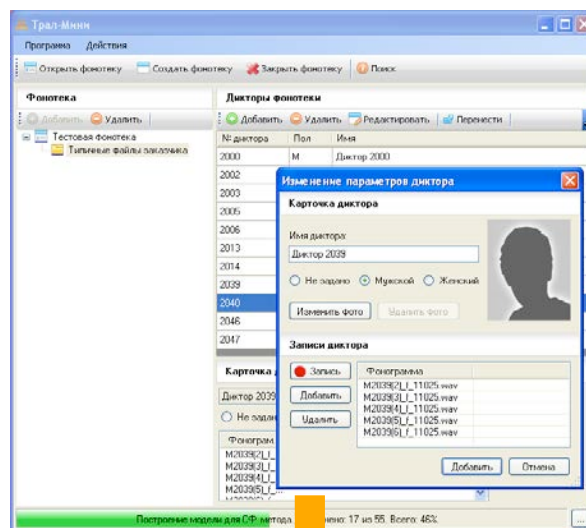


## Customization - ability to adapt the system to the key parameters of search



Speech Database

Adaptation of parameters – taking features of a specific speech database into account



Identification results

- ➔ **TrawlLab** - Facilitating voice ID analysis while carrying out multi-target forensic investigation by eliminating imposters and ranging the top-in-the-list speakers according to likelihood probability.

The screenshot displays the TrawlLab software interface. The left sidebar shows a file explorer with folders like 'Фонотека', 'аталоны', 'спорные', 'раздел1', and 'раздел2'. The main window has a menu bar with 'Программа', 'Действия', and 'Помощь'. Below the menu bar are buttons for 'Создать фонотеку', 'Открыть фонотеку', 'Закрыть фонотеку', and 'Поиск'. The main area is divided into two sections: 'Исходные дикторы' (Source dictators) and 'Результаты поиска' (Search results).

**Исходные дикторы (Source dictators):**

№ диктора	№ фонограммы	Пол	Имя диктора
0001	0001	М	M2000(4)_f_11025
0001	0002	М	M2000(4)_f_11025
0002	0001	М	M2002(5)_f_11025
0003	0001	М	M2328(4)_f_11025
0003	0002	М	M2328(4)_f_11025
0004	0001	М	M2329(5)_f_11025
0005	0001	М	M2337(4)_f_11025
0005	0002	М	M2337(4)_f_11025
0006	0001	М	M3004(5)_f_11025
3017	0001	М	M3017(4)_f_11025

**Результаты поиска (Search results):**

№ диктора	№ фонограммы	Пол	Имя диктора	SF FR	SF FA	PS FR	PS FA	GMMSVM FR	GMMSVM FA	Сходство
2000	0001	М	M2000(3)_f_11025	92.99	0.02	52.7066	1.22421	72.2918	0.093306	66.4534
2328	0001	М	M2328(2)_f_11025	0.09	54.29	0.2	62.7943	0.0788884	98.9712	0.0116506
2005	0001	М	M2005(1)_f_11025	34.11	2.82	2.09165	21.726	3.2924	24.9616	0.232501
2329	0001	М	M2329(3)_f_11025	21.86	4.84	0.247695	41.0464	2.9264	28.3722	0.167945
2337	0001	М	M2337(2)_f_11025	23.27	4.59	0.385303	36.091	0.272997	92.8958	0.0233982
3004	0001	М	M3004(1)_f_11025	0.37	35.82	0.2	70.5161	0.543534	81.3515	0.0197993
2002	0001	М	M2002(2)_f_11025	0	67.38	0.261456	39.3285	0.414599	86.9103	0.0164155
2005	0002	М	M2005(1)_f_11025	0.09	38.25	1.10491	28.4909	0.630067	77.7405	0.0224116
2000	0002	М	M2000(3)_f_11025	97.38	0	0.276921	38.9311	79.3615	0.0561777	55.4897
2329	0002	М	M2329(3)_f_11025	4.95	12.99	0.2	45.5236	1.28192	55.8527	0.0483441

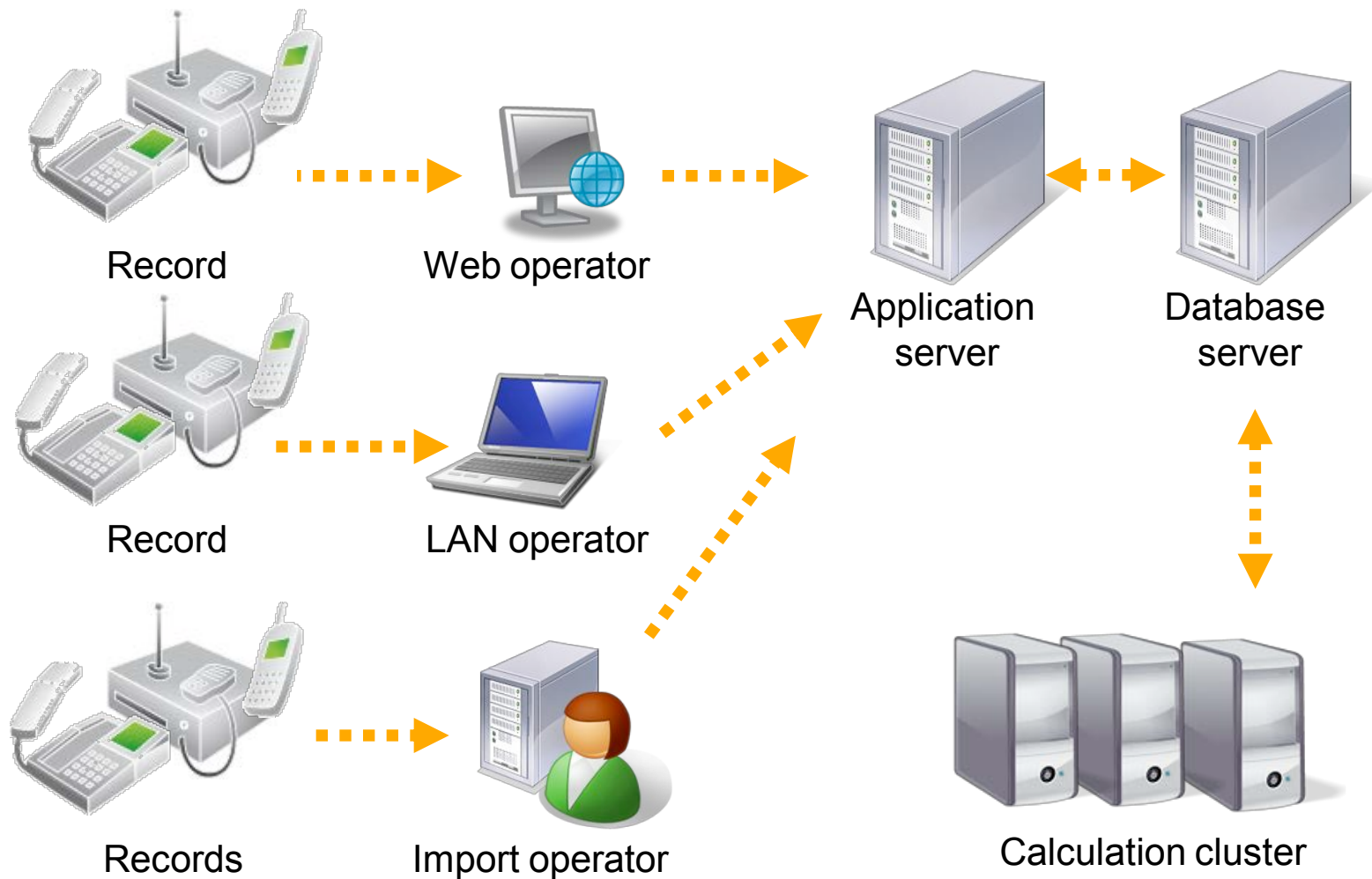
## VoiceNet.ID is designed for:

Reliable identification on a nation-wide voice database of speakers.

## VoiceNet.ID highlights

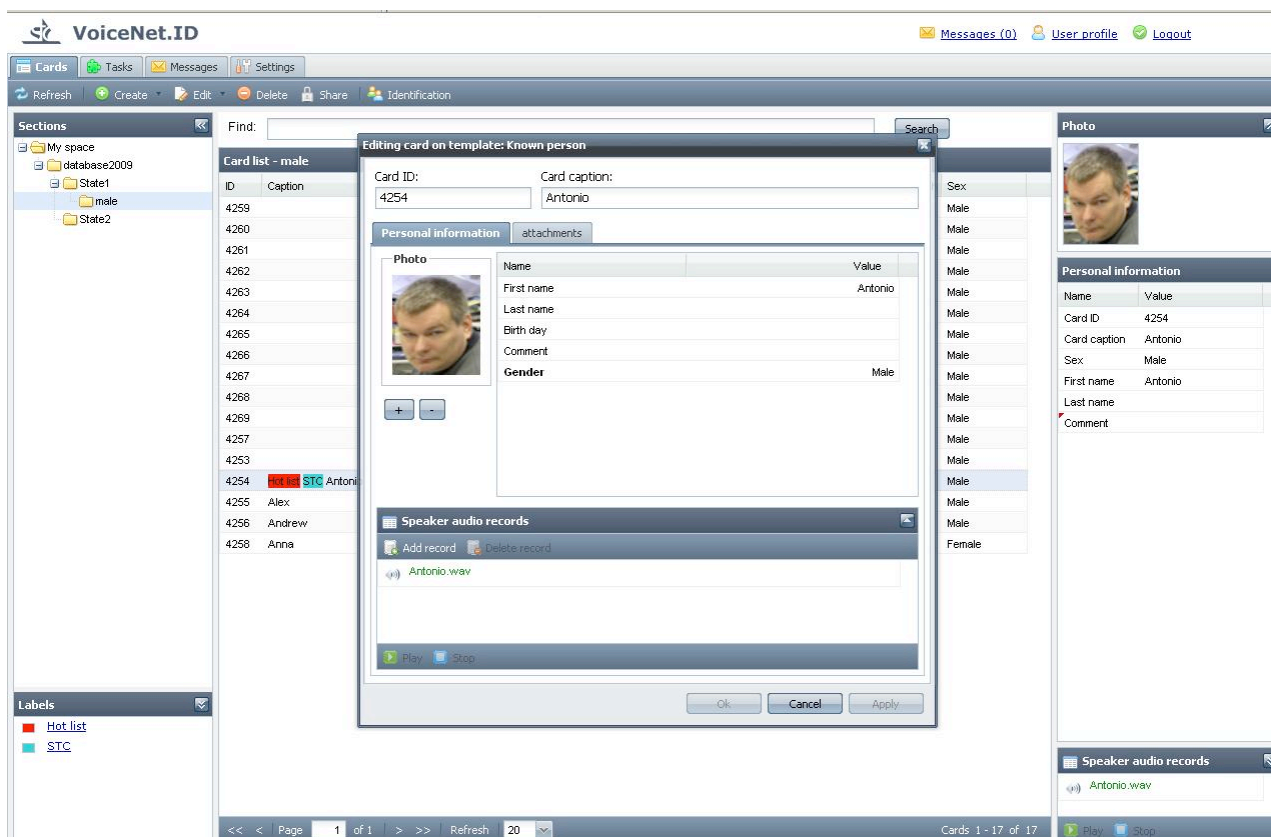
- ➔ Storage and real-time processing of large volume of voiceprints
- ➔ Client-server architecture
- ➔ Web-client
- ➔ Centralized speakers' profiles repository
- ➔ Multi-user system
- ➔ Secure storage and access
- ➔ Remote access to the database
- ➔ Additional information storage (video, photo, text)

## Architecture



## Speaker's profile card (SPC)

Automatically extracts biometric traits of voice and speech from the attached sound records. Speaker card can contain wealth additional information about the person (text, photo, video etc).



The screenshot displays the VoiceNet.ID web application interface. The main window shows a list of speaker cards on the left, with a central modal window for editing a specific card. The modal window is titled "Editing card on template: Known person" and contains fields for Card ID (4254) and Card caption (Antonio). Below these are tabs for "Personal information" and "attachments". The "Personal information" tab is active, showing a photo of a man and a table of personal data:

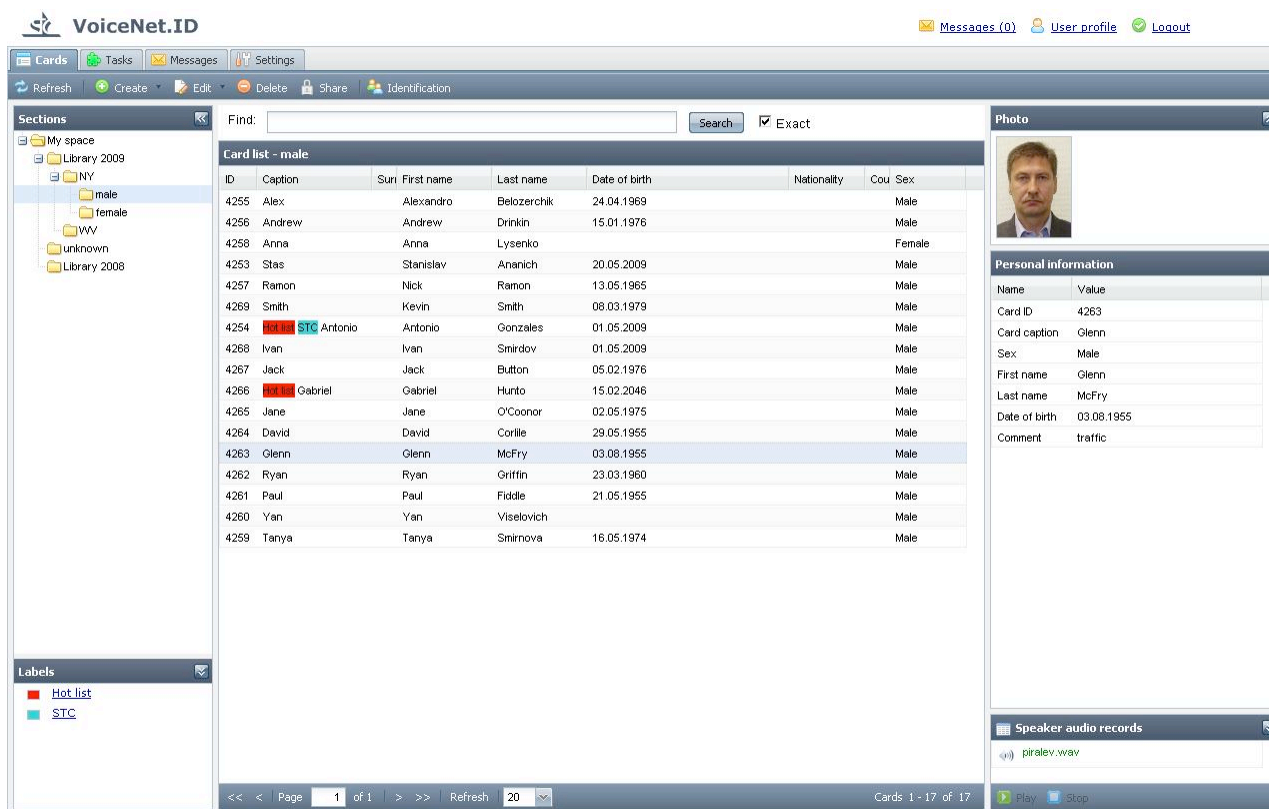
Name	Value
First name	Antonio
Last name	
Birth day	
Comment	
Gender	Male

Below the personal information is a section for "Speaker audio records" with buttons for "Add record" and "Delete record". A list of audio records is shown, including "Antonio.wav". At the bottom of the modal are "Ok", "Cancel", and "Apply" buttons.

The background interface includes a sidebar with "Sections" (My space, database2009, State1, male, State2) and "Labels" (Hot list, STC). The top navigation bar includes "Cards", "Tasks", "Messages", and "Settings". The bottom status bar shows "Page 1 of 1" and "Cards 1 - 17 of 17".

## Database management

SPCs in the database can be organized into unlimited number of sections and sub-sections to facilitate further search.



The screenshot displays the VoiceNet.ID web application interface. The top navigation bar includes links for Messages (0), User profile, and Logout. Below this, a toolbar offers actions like Cards, Tasks, Messages, and Settings. The main interface is divided into three primary sections:

- Sections:** A tree view on the left showing a hierarchy of folders: My space, Library 2009, NY, male, female, WW, unknown, and Library 2008.
- Card list - male:** A central table displaying a list of speaker cards. The table has columns for ID, Caption, Sun, First name, Last name, Date of birth, Nationality, Cou, and Sex. The card for Glenn McFry (ID 4263) is highlighted.
- Personal information:** A sidebar on the right showing details for the selected card, including a photo, name, card ID, card caption, sex, first name, last name, date of birth, and comment.

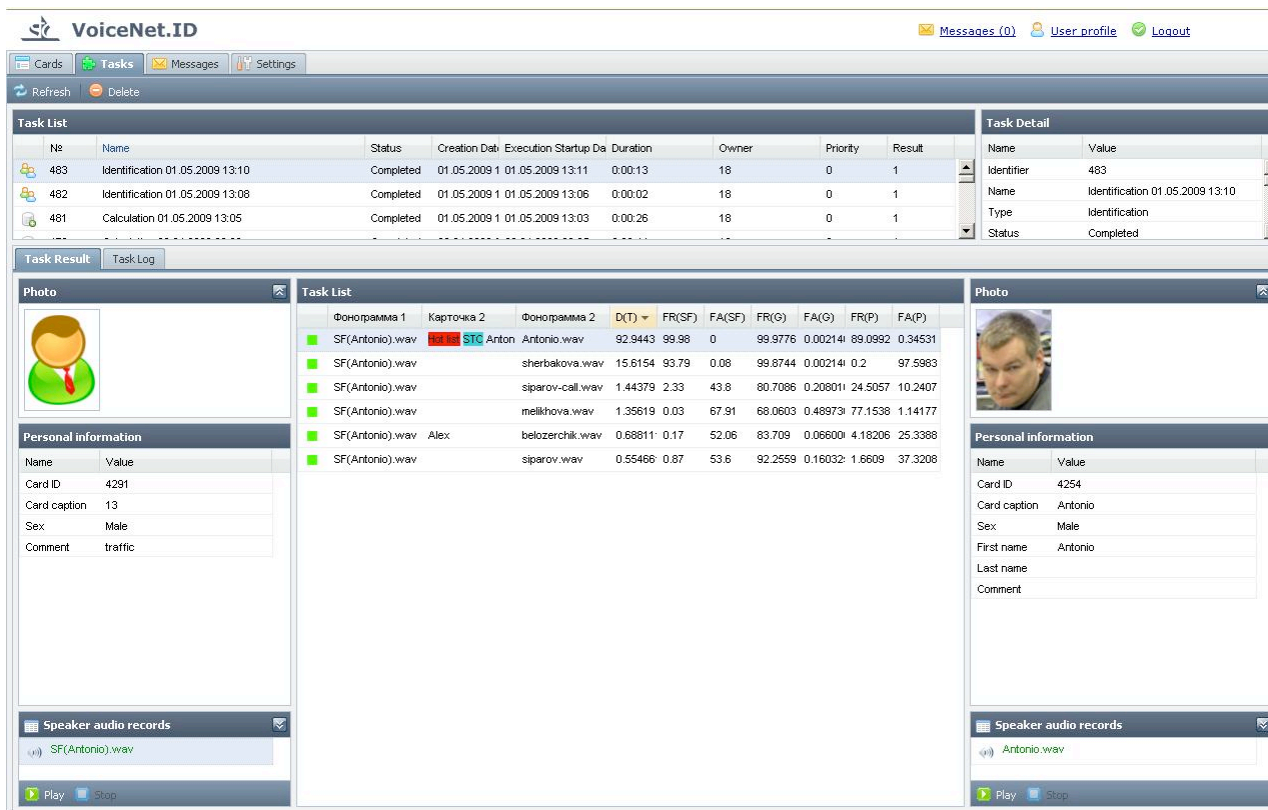
At the bottom, a status bar indicates the current page (1 of 1) and the total number of cards (17 of 17). A 'Speaker audio records' section at the bottom right shows a file named 'pirelev.wav' with play and stop controls.

ID	Caption	Sun	First name	Last name	Date of birth	Nationality	Cou	Sex
4255	Alex		Alexandro	Belozerschik	24.04.1969			Male
4256	Andrew		Andrew	Drinkin	15.01.1976			Male
4258	Anna		Anna	Lysenko				Female
4253	Stas		Stanislav	Ananich	20.05.2009			Male
4257	Ramon		Nick	Ramon	13.05.1965			Male
4269	Smith		Kevin	Smith	08.03.1979			Male
4254	Hot list STC Antonio		Antonio	Gonzales	01.05.2009			Male
4268	Ivan		Ivan	Smirdov	01.05.2009			Male
4267	Jack		Jack	Button	05.02.1976			Male
4266	Hot list Gabriel		Gabriel	Hunto	15.02.2046			Male
4265	Jane		Jane	O'Connor	02.05.1975			Male
4264	David		David	Corlie	29.05.1955			Male
4263	Glenn		Glenn	McFry	03.08.1955			Male
4262	Ryan		Ryan	Griffin	23.03.1960			Male
4261	Paul		Paul	Fiddle	21.05.1955			Male
4260	Yan		Yan	Viselovich				Male
4259	Tanya		Tanya	Smirnova	16.05.1974			Male



## Identification results

The results of “**VoiceNet.ID**” search presented in the form of a list with indication of likelihood probability (LR) of each record containing the speech of a target speaker.



The screenshot displays the VoiceNet.ID web application interface. At the top, there is a navigation bar with links for Messages (0), User profile, and Logout. Below this, a sidebar contains links for Cards, Tasks, Messages, and Settings. The main content area is divided into two sections: Task List and Task Detail.

**Task List**


No	Name	Status	Creation Date	Execution Start Date	Duration	Owner	Priority	Result
483	Identification 01.05.2009 13:10	Completed	01.05.2009 1	01.05.2009 13:11	0:00:13	18	0	1
482	Identification 01.05.2009 13:08	Completed	01.05.2009 1	01.05.2009 13:06	0:00:02	18	0	1
481	Calculation 01.05.2009 13:05	Completed	01.05.2009 1	01.05.2009 13:03	0:00:26	18	0	1

**Task Detail**

Name	Value
Identifier	483
Name	Identification 01.05.2009 13:10
Type	Identification
Status	Completed

**Task Result** | **Task Log**

**Photo**



**Personal information**

Name	Value
Card ID	4291
Card caption	13
Sex	Male
Comment	traffic

**Speaker audio records**

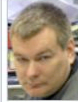
SF(Antonio).wav

Play Stop

**Task List**

Фонотрама 1	Карточка 2	Фонотрама 2	D(T)	FR(SF)	FA(SF)	FR(G)	FA(G)	FR(P)	FA(P)
SF(Antonio).wav	STC Anton	Antonio.wav	92.9443	99.98	0	99.9776	0.00214	89.0992	0.34531
SF(Antonio).wav		sherbakova.wav	15.6154	93.79	0.08	99.8744	0.00214	0.2	97.5983
SF(Antonio).wav		siparov-call.wav	1.44379	2.33	43.8	80.7086	0.20801	24.5057	10.2407
SF(Antonio).wav		melikhova.wav	1.35819	0.03	67.91	68.0603	0.48973	77.1538	1.14177
SF(Antonio).wav	Alex	belozherchik.wav	0.68811	0.17	52.06	83.709	0.06600	4.18206	25.3388
SF(Antonio).wav		siparov.wav	0.55466	0.87	53.6	92.2559	0.16032	1.6609	37.3208

**Photo**



**Personal information**

Name	Value
Card ID	4254
Card caption	Antonio
Sex	Male
First name	Antonio
Last name	
Comment	

**Speaker audio records**

Antonio.wav

Play Stop

## Technical specs:

- ➡ DBMS - Oracle 11g, PostgreSQL, ready to be adapted for others
- ➡ OS – UNIX (Solaris 10, Linux), Windows Server 2003 or later
- ➡ Web Service based architecture
- ➡ Application Server (GlassFish V3, Tomcat 6, ready to be adapted for others )
- ➡ Cluster calculations JPPF 1.8

## Performance & scalability:

- ➡ Size – Database is scalable up to 10`000`000 cards
- ➡ Speed – Performance directly linked to the computing power of a server (parallel calculation support)
- ➡ Tasks – The system can be adopted to any voice ID challenge (search for unknown speakers in the database or search for known speakers in the stream of audio files)

Thank you for your attention!

[WWW.SPEECHPRO.COM](http://WWW.SPEECHPRO.COM)

tel.: +7 812 331-0665

fax: +7 812 327-9297

