



Voice Biometrics
Speech Recognition
Audio Watermarking

info@dtec-bio.es
www.dtec-bio.es



DTec: R&D in Speech Technologies

- ✓ Technology fully created from **in house R&D**.
- ✓ Available both **on-premise** for several OSs and as a cloud web service for a **SaaS** alternative.
- ✓ **BioVox**: advanced security through voice biometrics. *Who?*
- ✓ **ReconVox**: speaker independent continuous speech recognition. *What?*
- ✓ **AudioWatermark**: embedding of hidden, non audible information into an audio stream.





BioVox

Voice Biometrics advanced security

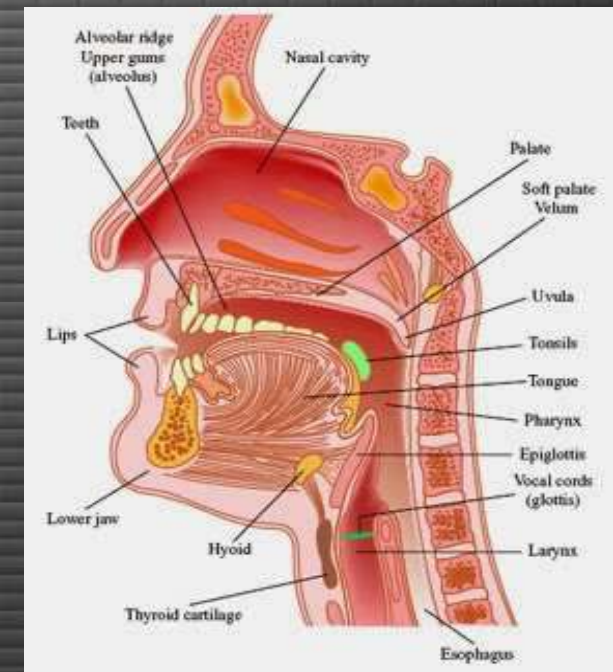
- ✓ **Identification** of speakers using their **voice**: *Who?*
- ✓ **Text** and language **independent**.
- ✓ Both speaker **verification** (1:1) and **identification** (1:N) tasks.
- ✓ **High accuracy**: EER < 1%.
- ✓ Extremely **efficient** verification engine:
 - ✓ **High speed**: verification response time < 0.5 s.
 - ✓ **Low memory requirements**: can be integrated into embedded systems.

Voice Biometrics: *BioVox*

Principles



- ✓ Based on **physical features** of the human vocal tract, **unique** for each individual.
- ✓ More than **1.000 parameters** are modeled and stored into a **voiceprint**.
- ✓ Physical features **can't be impersonated**.
- ✓ **Spoofing attacks** using voice recordings are **countered** requesting random **captchas** with *ReconVox*.





Voice Biometrics: *BioVox* *Applications*



- ✓ **Security** in phone banking and e-commerce.
- ✓ Speaker **authentication** in the background during all the phone call.

- ✓ **Contactless physical access control and check-in** for employees:
 - Hygienic.
 - Resistant to blows and vandalism.





Voice Biometrics: *BioVox* *Applications*



✓ Remote presence control: GPS geolocation + speaker verification.



✓ Police and forensic investigations.

✓ Biometric security in embedded systems.



ReconVox

Speaker independent continuous speech recognition



- ✓ Simple **voice commands** or **transcription of a long sentence**:
What?
- ✓ **Speaker independent**: no need for custom training.
- ✓ **AutoLearn**: automatic accuracy improvement. It can adapt to a specific speaker, dialectic region or noisy acoustic channel.
- ✓ **Word-spotting**: keywords detection. Independent of the language of the speech surrounding the word. No grammars or predefined dictionaries are needed.



Speech Recognition: *ReconVox* *Applications*



- ✓ *IVRs* in call-centers.
- ✓ *Automatic transcription* of continuous speech.



- ✓ *Media clipping*.
- ✓ *Speech Analytics*: keywords search into voice recordings.



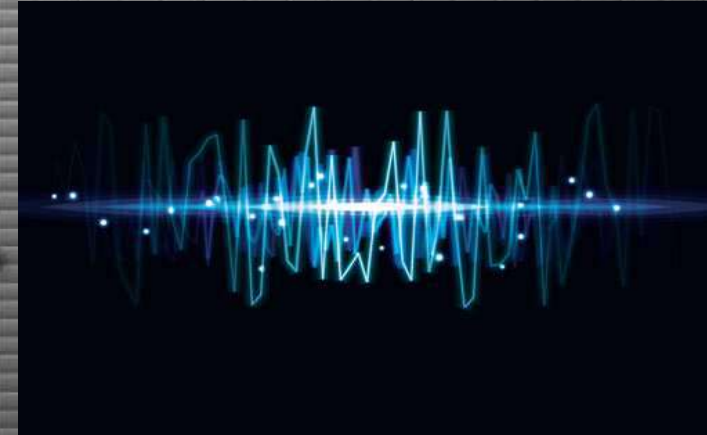
Speech Recognition: *ReconVox* *Applications*



- ✓ Together with *BioVox*, **secure** voice commands:
 - Responding just to the **authorized speaker**.
 - **Domotics** and alarm control.
 - Electronic devices in **help** for the **impaired people**.

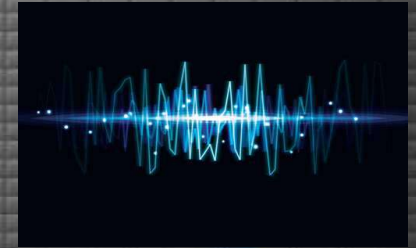


AudioWatermark



- ✓ **Steganography**: embedding of **hidden**, not hearable **information** into an audio stream.
- ✓ Audio watermark **attack resistant**:
 - Compression.
 - Recoding.
 - D/A/D conversion: **live microphone recording**.
- ✓ Audio **integrity** guaranteed: cut and paste attacks are detected.

AudioWatermark Applications



- ✓ **Tracking** of audio recordings from call-centers.
- ✓ **Copyright** control for music files.
- ✓ **Detection of manipulated audio** (insertion or deletion of segments) in sensitive conversation recordings.
- ✓ **Transmission of hidden information** in radio broadcast or phone calls.





DTEC

info@dtec-bio.es
www.dtec-bio.es



AudioWatermark is *steganography* technology developed by **DTec**. With **AudioWatermark** you can embed **hidden information into audio signals** (both live streams and recordings) and recover it later after transmission. The message is actually merged with the audio signal, totally different to a simple inclusion of hidden bits into a digital audio file or stream.

The information is **inaudible**, it can't be extracted by unauthorized listeners and it's **robust** to the most usual channel distortions and compressions, like MP3. It can even survive Digital to Analog and back to Digital conversions.

In addition to secret info transmission, because the watermark can't be added, removed or modified in any way by attackers without affecting it, it's also possible to guarantee the **integrity of the original signal**, detecting manipulations like cut and paste of audio segments.

AudioWatermark is an open **SDK (Software Development Kit)** that exports its functionalities through an easy to use API (*Application Programming Interface*). With this API you can start marking audio from any embedded hardware or software platform.

The complete watermarking communication scenario can be seen as **three successive stages**:

1. **Watermark embedding**: some secret info is merged with the carrier audio signal while keeping the acoustic properties unaffected for the Human Hearing System.
2. **Data transmission**: the audio signal together with the hidden info is transmitted through a communication channel. This channel is an abstraction for all the degradations and **attacks** the watermark is going to suffer before its later extraction and can consist of physical transmission through a noisy audio channel, digital to analog conversions, on air recording, resampling, recoding, etc.
3. **Watermark extraction**: the audio signal is processed by an analysis module that rebuilds and extracts the hidden information embedded in the first stage.

The capabilities of **AudioWatermark** make it invaluable in many different situations:



- **Keeping track of the identity** of the user that retrieved a specific recording from a call-center.
- Sending **secret information** hidden into radio or TV broadcasts.
- Enforcing **copyright in music** audio files.
- Guaranteeing the **integrity of important voice recordings**, like contracts completed by phone or recorded sessions in trials.

PRODUCT

- Robust and secure embedding of hidden information into an audio signal.

KEY FEATURES



- Information is **hidden into the actual audio signal**, not at file or digital stream level.
 - Watermarked audio **can't be distinguished from the original** (based on Human Hearing System properties).
 - Information **can't be read, removed or modified** by attackers.
 - **Cut and paste attacks** in the watermarked audio **are detected**.
 - Robust to **Digital / Analog / Digital** conversions.
 - Robust to **MP3 recoding**.
- Robust to **frequency resampling**.
 - Robust to **noisy phone channels**.

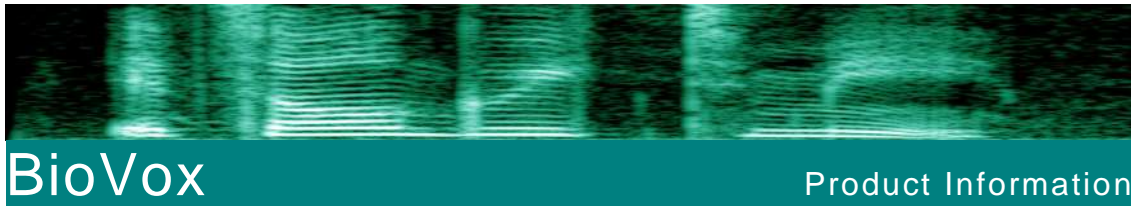
TECHNICAL SPECIFICATIONS

- Length of hidden information can be configured; minimum **8 bits**.
- Absolute minimum audio for watermarking: **2.4 seconds** (for minimum watermark length).
- Recommended minimum audio for watermarking: **7.2 seconds** (for minimum watermark length).
- Watermark embedding speed¹: **100X faster than real time**.
- Supported formats for carrier audio: PCM linear 16 bits 8/16 KHZ, A-Law, μ -Law.
- Minimum recommended CPU: Intel i3 @ 2'5 GHz.

SUPPORTED PLATFORMS

- Windows® XP, Vista, 7, 8, 10.
- Linux, several distributions.

¹ With minimum recommended CPU.



BioVox is our speaker verification and identification product. Thanks to its advanced **voice biometrics** technology you can raise the security level in physical or logical access controls and at the same time get rid of magnetic cards and passwords. So you have a double gain: **increased security + user friendly systems**. You can also identify speakers in real time during a phone call.

Because it's **text independent**, **BioVox** is in a different level compared to many other voice biometrics solutions. Users don't need to say fixed pass phrases like "my voice is my passport", so **spoofing attacks are almost completely eradicated** because a different random sequence of numbers can be requested in each login. On the other hand, this ability grants a huge flexibility because **voiceprints can be generated from past free speech recordings**.

BioVox provides an open **SDK (Software Development Kit)** that exports its functionalities through a powerful yet easy to use API (*Application Programming Interface*). With this API you can integrate a complete voice based user validation system into any embedded hardware or software application.

The validation process is done in **two successive steps**, enrollment and validation:

- **Enrollment:** the new user pronounces several utterances, which are then analyzed in order to extract a **voiceprint** that identifies the speaker in a unique way.
- **Validation:** the user to be validated pronounces an utterance (free text, his/her name or a password) which is analyzed and compared with the associated voiceprint, if we're in a speaker **verification** scheme. If they match, the legitimate user is accepted. In the other hand, if the application is working in a speaker **identification** scheme, the utterance is compared with all the voiceprints available and returns the associated user identity, if a match is found.

The open architecture of **BioVox** makes possible a wide range of different applications:

- Security in **call-centers**: continuous identity verification performed in the background.
- **e-commerce & e-banking**: secure payment in Internet, payment with the mobile phone.
- Physical **access controls** and presence controls: no more buddy punching.
- **Alarms and domotics**: electronic devices driven by secure voice commands.
- **Police investigations**, forensic acoustics: identification of suspects in real time.
- **Search for specific speakers in audio recordings**.



PRODUCT

- Text Independent Speaker Verification and Identification System.

KEY FEATURES



- Two matching modes: **verification** (1:1 matching) and **identification** (1:N matching).
- **Text independent.**
- **Language independent.**
- **Security level** can be adjusted.
- Automatic calculation of **quality of voiceprints.**
- Automatic calculation of the **matching score** between the speaker and the voiceprint.
- Two operation modes: **real time** or **batch** mode (file based).
- Highly optimized verification engine: can be integrated into **embedded systems.**

TECHNICAL SPECIFICATIONS

- Audio for enrollment: 30 s minimum, >60 s recommended.
- Audio for validation: 2 s minimum, >5 s recommended.
- Supported audio formats: PCM linear 16 bits 8/16 KHZ, G.711, MP3 (not recommended).
- Voiceprint size: 4 KB.
- Verification (1:1) time¹: < 0.4 seconds.
- Identification (1:N) rate¹: 100 voiceprints analyzed / second.
- EER²: < 1%, dependent of application and system configuration.
- Minimum recommended CPU: Intel i3 3GHz or equivalent.

SUPPORTED PLATFORMS

- Windows® XP, Vista, 7, 8, 10.
- Linux, several distributions.
- Android® NDK.

¹ With minimum recommended CPU.

² *Equal Error Rate*: the value in which the two opposite error rates associated to any biometric system are equal after tuning (whenever one is reduced, the other one is increased as a consequence). These error rates are: *FRR* (False Rejection Rate - a legitimate user is rejected) and *FAR* (False Acceptance Rate - an impostor is wrongly accepted).



ReconVox is our high performance **Speaker Independent Continuous Speech Recognition** product. Thanks to its ability to work both in **Word-Spotting** or continuous speech mode for any speaker without needing specific training, it fits into a wide range of applications, from controlling electronic devices with your voice to accessing through the telephone line automatic services driven by full and complex sentences or getting the full transcription of a radio broadcast.

ReconVox is designed to be easily integrated in almost any operating environment because it's available as a **SDK (Software Development Kit)** which delivers all its functionality via a complete **API (Application Programming Interface)**.

Thanks to its efficiency and flexibility, a wide range of **applications** can be delivered:

- **IVR (Interactive Voice Response):** conversations close to natural language in automatic call centers.
- **Automatic search by content:** spotting of **keywords** or sentences in audio/video recordings or streaming audio.
- **Automatic transcription** services from radio broadcasts, trials...
- **Alarms and domotics:** electronic devices controlled by voice commands, (de)activation of alarms...
- **Voice commands in cars:** GPS, hands-free phone calls...



- **Media clipping** and **automatic tagging of contents.**
- **Education:** scoring of **pronunciation** for every single word in language learning or in some speech pathologies like dyslexia or aphasia.

In addition, if security is a factor for the application, a **ReconVox** based Speech Recognition system can work together with our Voice Biometrics technology, **BioVox**. This way, it's possible to perform a continuous speaker verification along all the interaction of the user with the system, in the background and transparent for him or her, for example for secure transactions in financial call centres.

PRODUCT

- Speaker Independent Continuous Speech Recognition system.

KEY FEATURES

- Recognition task can be fine tuned: **isolated words** or **continuous speech**.
- **Speaker independent**: doesn't need to be retrained for every speaker.
- **AutoLearn**: automatic adaptation for a specific speaker, dialectic region or noisy environment.
- **WordSpotting**: detection of keywords or special sentences among out of vocabulary words.
- **Vocabulary can be customized**: from a few commands to thousands of words.
- **Grammar and language model can be customized**: two different types of language models, **fixed syntax** or **flexible (statistic) grammar**.
- Available in **spanish** and **UK english**. New languages can be incorporated upon request.
- Acoustic models available in **8 KHz** (telephone) and **16 KHz** (radio, domotics, apps).
- **Efficient** recognition engine: can be integrated into embedded systems.



TECHNICAL SPECIFICATIONS

- Speech signal preprocessing: automatic activity detection and signal filtering.
- Recognition speed in continuous speech mode¹: 2x – 4x faster than real time (typical, depends of vocabulary size).
- Recognition speed in Word-Spotting mode¹: 7x faster than real time.
- Supported audio formats: PCM linear 16 bits 8/16 KHz, A-Law, μ -Law, MP3.
- Memory requirements: 3 MB fixed + 9 MB language models + 20 MB *AutoLearn* (optional, can be increased).
- Disk space: 5 MB / language.
- Minimum recommended CPU: Pentium IV 2'5 GHz.

SUPPORTED PLATFORMS

- Windows® XP, Vista, 7, 8, 10.
- Linux, several distributions.

¹ With minimum recommended CPU.