

# شناسا

سامانه شناسایی گوینده

از روی صدا

\* تایید هویت گوینده

\* تشخیص هویت گوینده



گفتار اگر چه در ظاهر، شامل پیامی از کلمات پشت سر هم است، اما در کنار آن شامل اطلاعاتی در مورد نوع زبان، وضعیت احساسی گوینده، جنسیت و به طور کلی هویت فرد می‌باشد. هدف شناسایی گوینده، استخراج اطلاعاتی از سیگنال گفتار است که در بردارنده هویت منحصر به فرد گوینده می‌باشد. شناسایی گوینده شامل دو حوزه زیر است:

- تایید هویت از روی صدا
  - تشخیص هویت گوینده از روی صدا
- در تایید هویت ادعای فرد مبنی بر اینکه چه کسی است، تایید یا رد می‌شود و در تشخیص یا تعیین هویت گوینده، فرد از روی صحبت کردن شناسایی می‌شود.
- شناسا، سامانه شناسایی گوینده از روی صدا است که می‌تواند با هدف تایید یا تشخیص هویت در کاربردهای مختلف امنیتی و کنترل دسترسی، به تنهایی یا در کنار روش‌های امنیتی دیگر مورد استفاده قرار گیرد. شناسایی گوینده از روی صدا با توجه به اینکه صدای فرد همواره همراه وی بوده و معایبی مانند گم شدن را ندارد و می‌تواند بدون حضور فیزیکی و از راه دور (مانند پشت تلفن) مورد استفاده قرار گیرد، به سایر روش‌ها مزیت دارد.

## کاربردهای سامانه شناسایی گوینده شناسا

- کنترل دسترسی از راه دور و استفاده از گفتار به عنوان یک رمز عبور
- جستجو و شناسایی مجرمین از روی صدا
- تشخیص گوینده و دسته‌بندی مکالمات
- نظارت بر مکالمات و گفتگوها و اقدام جهت آرشيو یا اعلام خبر و امثال آن



# سامانه تایید هویت از روی صدا - شرح فنی

به طور کلی این سامانه به دو مرحله مجزا تقسیم می‌شود. مرحله‌ی ثبت‌نام و مرحله‌ی تایید هویت

## ۱) مرحله‌ی ثبت نام

در این مرحله، برای هر گوینده به صورت خودکار و با استفاده از فایل‌های از قبل ضبط شده، یک مدل ساخته می‌شود. برای این کار، ابتدا از گفتارهای ورودی ویژگی‌هایی استخراج می‌شود، سپس با استفاده از روش‌های مختلف مدل‌سازی، ویژگی هر گوینده مدل می‌شود. در حالت مستقل از متن، حداقل گفتار مورد نیاز برای آموزش مدل هر گوینده ۳۰ ثانیه است. در حالت‌های وابسته به متن و استفاده از متن تصادفی، حداقل سه بار تکرار عبارت مورد نظر لازم است. برای بالا رفتن دقت تایید هویت، بهتر است از یک گوینده در چندین نشست صدای ضبط شده وجود داشته باشد. مدل‌های ساخته شده برای گویندگان در پایگاه داده‌ای ذخیره می‌شوند. در نهایت برای انجام عمل تایید هویت نیاز است که برای هر گوینده آستانه‌ای محاسبه شود. این آستانه بعد از ساخته شدن مدل هر گوینده محاسبه می‌شود.

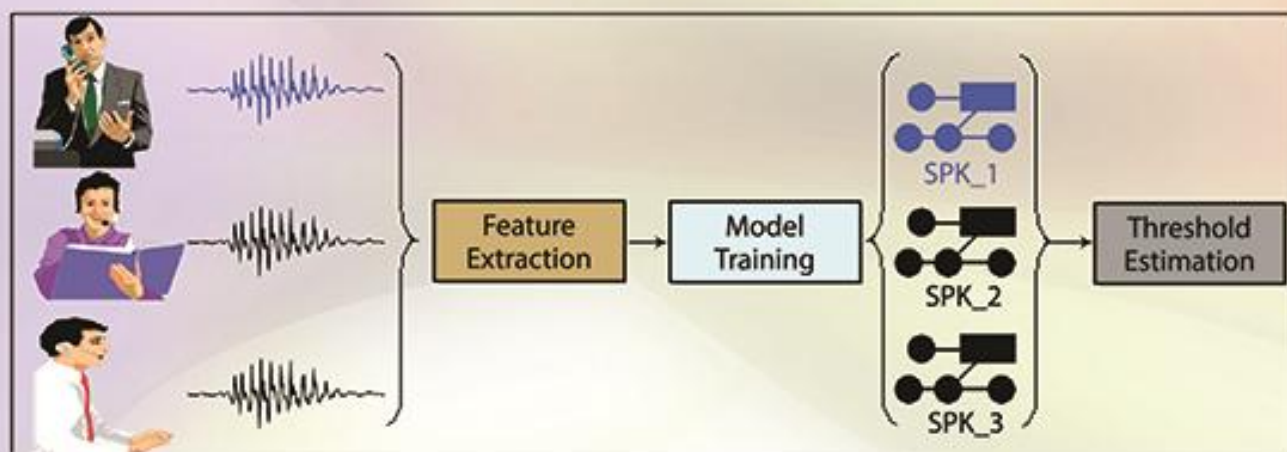
وظایف هر کدام از ماژول‌های مرحله ثبت‌نام به صورت زیر است:

**Feature Extraction:** این بخش وظیفه استخراج ویژگی‌های سیگنال گفتار گوینده‌ها را بر عهده دارد.

**Model Training:** این واحد آموزش صدای گوینده‌ها را بر عهده دارد. در این بخش از مدل‌سازی آماری استفاده می‌شود.

**Threshold Estimation:** در این ماژول

برای گوینده‌ها مقادیری به عنوان سطح آستانه تخمین زده می‌شود که بر اساس تشخیص هویت صورت می‌گیرد.

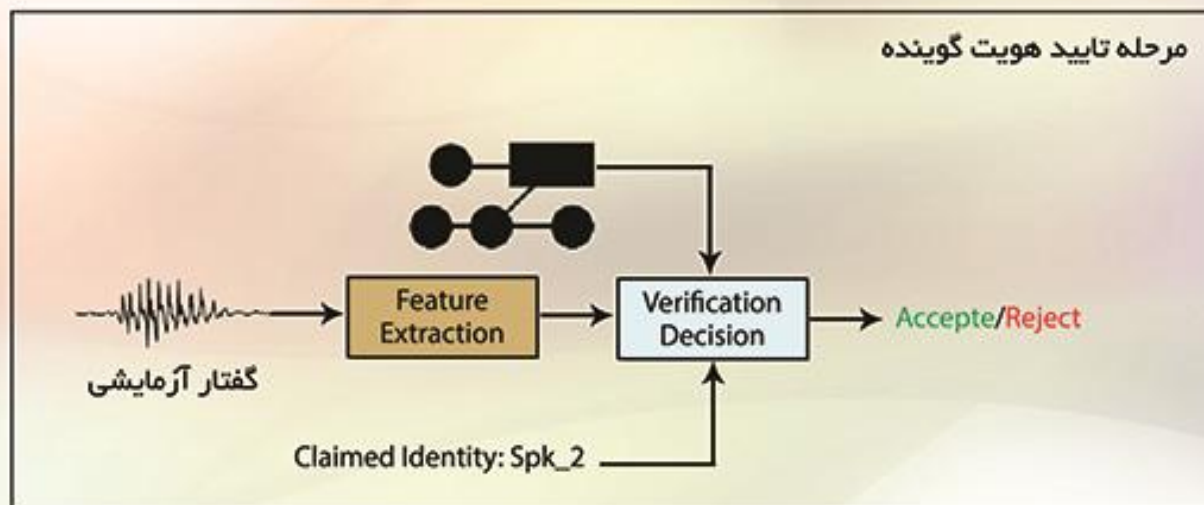


## ۲) مرحله‌ی تایید

در مرحله‌ی تایید، ابتدا سیستم امتیاز گفتار ورودی را با استفاده از مدل گوینده‌ی مورد ادعا که از قبل ساخته شده است، بدست می‌آورد. سپس این امتیاز با آستانه‌ی بدست آمده در مرحله‌ی آموزش مقایسه می‌شود و در صورتی که بیشتر از آن باشد گوینده تصدیق می‌شود. حداقل گفتار ورودی برای انجام عمل تایید وابسته به روش استفاده شده است. در حالتی که از روش مستقل از متن استفاده شود این مقدار متغیر است و حداقل آن شش ثانیه است. برای روش استفاده از گذرواژه‌ی ثابت حداقل یک بار تکرار گذرواژه لازم است. برای روش متن تصادفی، یک بار تکرار عبارتی با ۴ کلمه کافی است. گام‌های این مرحله در شکل زیر نشان داده شده است.

وظایف هر کدام از ماژول‌های این بخش به صورت زیر است:

- **Feature Extraction**: این بخش که با بخش استخراج ویژگی آموزش یکسان است، وظیفه استخراج ویژگی‌های سیگنال گفتار گوینده‌ها را بر عهده دارد.
- **Verification Decision**: این ماژول بر اساس مقایسه امتیاز صدای ورودی با مدل گوینده هدف، در مورد تایید یا رد تصمیم‌گیری می‌کند.



# شناسا - تشخیص: سامانه تشخیص هویت از روی صدا

شناسا-تشخیص، یک سامانه پردازش گفتار است که برای تشخیص افراد از روی صدا به کار می‌رود. این سامانه، در صدای دریافتی، از بین گویندگان مختلف به دنبال گوینده‌ی مورد نظر (هدف) می‌گردد. شناسا-تشخیص، مستقل از محتوای بیان شده، زبان و لهجه گوینده و همچنین مستقل از کانال انتقال گفتار، گوینده یک گفتار را شناسایی می‌کند. در صورتی که گفتار ورودی متعلق به هیچ کدام از گویندگان ثبت‌نام شده در سامانه نباشد، شناسا توانایی تشخیص آن را نیز دارد. اصلی‌ترین ویژگی‌های این سامانه عبارتند از:



- مستقل از زبان و لهجه گوینده
- مستقل از متن و محتوای مکالمه
- دقت بالا در تشخیص گوینده‌ها
- سرعت بالا در تشخیص گوینده‌ها به ویژه برای تعداد گوینده‌های بالا
- امکان گروه‌بندی گویندگان
- تشخیص گفتارهایی که متعلق به هیچ کدام از گویندگان نیست (مجموعه باز)
- قابلیت استفاده به صورت برخط (Online) و برون خط (Offline)
- پردازش همزمان کانال‌های تلفنی ورودی
- پشتیبانی از فرمت‌های مختلف صوتی و فایل‌های استریو
- مجهز به ویرایش‌گر صدا
- استفاده به صورت سرویس و ماژول
- امکان به‌کارگیری در سایر نرم‌افزارها و سیستم‌ها

# سامانه تشخیص هویت از روی صدا - شرح فنی

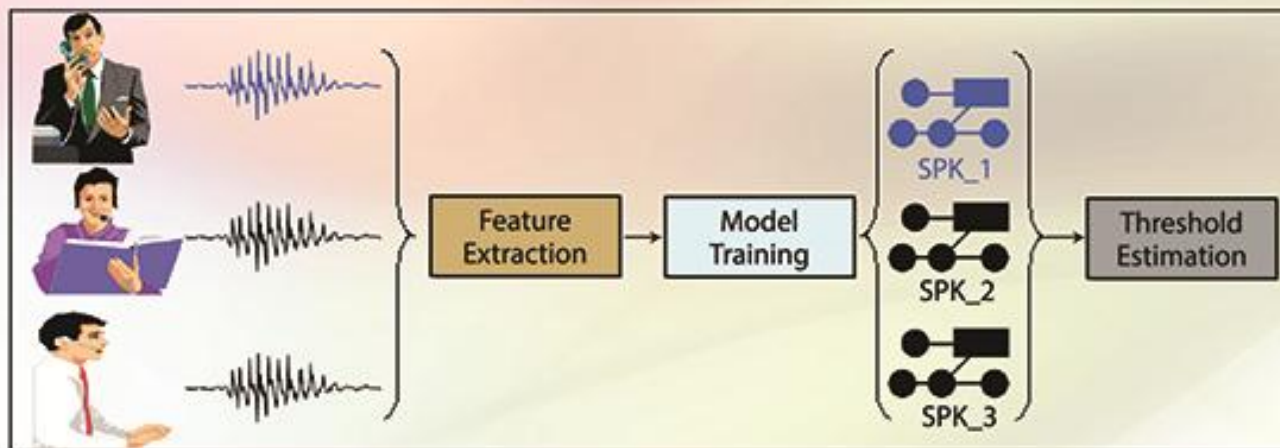
به طور کلی این سامانه به دو مرحله مجزا تقسیم می‌شود: مرحله‌ی ثبت‌نام و مرحله‌ی شناسایی.

## ۱) مرحله‌ی ثبت نام

در مرحله‌ی ثبت‌نام، برای هر گوینده به صورت خودکار و با استفاده از فایل‌های از قبل ضبط شده، یک مدل ساخته می‌شود. برای این کار، ابتدا از گفتارهای ورودی ویژگی‌هایی استخراج می‌شود. سپس با استفاده از روش‌های مختلف مدل‌سازی، ویژگی‌های هر گوینده مدل می‌شود. حداقل گفتار مورد نیاز برای آموزش مدل هر گوینده ۳۰ ثانیه است. برای بالا رفتن دقت شناسایی، بهتر است از یک گوینده چندین فایل از چندین کانال مختلف وجود داشته باشد. مدل‌های ساخته شده برای گویندگان در پایگاه داده‌ای ذخیره می‌شوند. در نهایت برای تشخیص گفتارهای خارج از گروه نیاز است که برای هر گوینده آستانه‌ای محاسبه شود. این آستانه‌ها بعد از ساخته شدن مدل تمام گویندگان محاسبه می‌شوند.

وظایف هر کدام از ماژول‌های فوق به صورت زیر است:

- **Feature Extraction**: این مرحله وظیفه استخراج ویژگی‌های سیگنال گفتار گوینده‌ها را بر عهده دارد.
- **Model Training**: این واحد آموزش صدای گوینده‌ها را بر عهده دارد. در این بخش از مدل‌سازی آماری استفاده می‌شود.



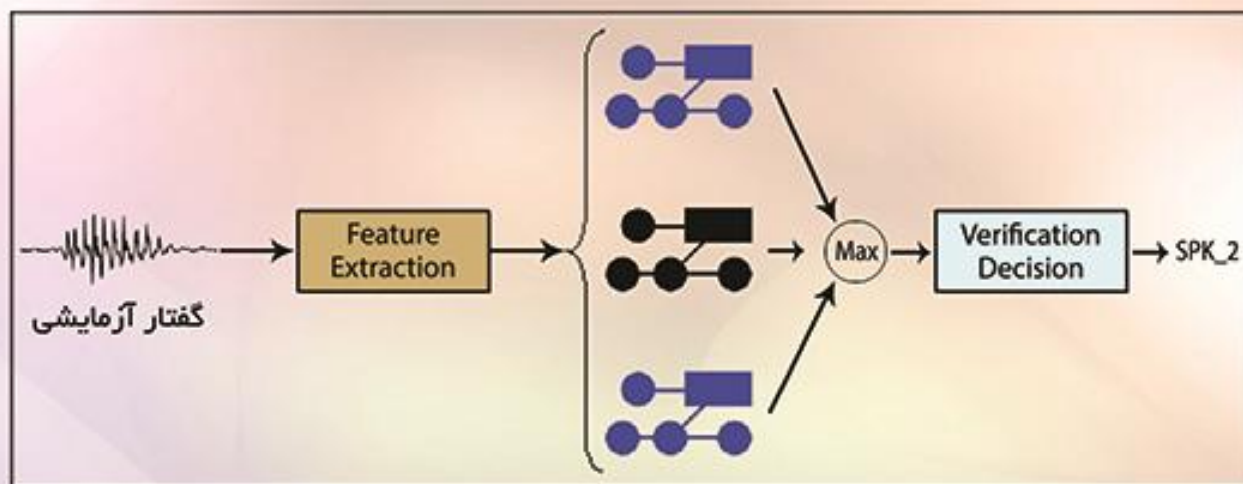
- **Threshold Estimation**: در این ماژول برای گوینده‌ها مقادیری به عنوان سطح آستانه تخمین زده می‌شود که بر اساس تشخیص هویت صورت می‌گیرد.

## مرحله شناسایی

در مرحله شناسایی، سیستم با استفاده از روش‌های مستقل از متن، گفتار ورودی را با مدل‌هایی که از قبل ساخته شده‌اند مقایسه می‌کند و گوینده هدف را در مدت زمان اندکی شناسایی می‌کند. حداقل گفتار ورودی برای انجام عمل شناسایی ۱۲ ثانیه است. نمای کلی این مرحله در شکل زیر نشان داده شده است.

وظایف هر کدام از ماژول‌های این بخش به صورت زیر است:

**Feature Extraction:** این بخش که با بخش استخراج ویژگی آموزش یکسان است، وظیفه استخراج ویژگی‌های سیگنال گفتار گوینده‌ها را بر عهده دارد.



**Decision Making:** این ماژول بر اساس امتیاز شباهت صدای ورودی به مدل‌های موجود در سیستم در مورد نتیجه خروجی تصمیم‌گیری می‌کند.

همان‌طور که اشاره شد سیستم توانایی تشخیص گفتارهایی که متعلق به هیچ کدام از گویندگان ثبت‌نامی نیست، را دارد. تشخیص این موضوع در قسمت تصمیم‌گیری انجام می‌شود. برای این کار، امتیازی که برای شبیه‌ترین گوینده به گفتار آزمایشی بدست آمده، با آستانه‌ای که در مرحله آموزش محاسبه شده مقایسه می‌شود. در صورتی که این مقدار بیشتر از آستانه بود گوینده به عنوان گوینده هدف شناسایی می‌شود و در غیر این صورت، گوینده به عنوان گوینده‌ی مظنون با درصد ظن مشخص، شناسایی می‌شود.





تلفکس: ۰۲۱) ۶۶ ۵۵ ۱۵ ۲۵  
[www.asr-gooyesh.com](http://www.asr-gooyesh.com)  
[info@asr-gooyesh.com](mailto:info@asr-gooyesh.com)