



Reconocimiento de voces
específicas con las API de
Speaker Recognition en
Azure Cognitive Services

BIENVENIDA

Obtener información sobre cómo usar Speaker Recognition API para identificar usuarios específicos a través de sus voces.

Objetivos de este módulo:

- Conocer el concepto de reconocimiento del hablante.
- Obtener información sobre los conceptos relacionados con el reconocimiento del hablante.
- Explorar las Speaker Recognition API.

2

Requisitos previos

Ninguno

Introducción 1 min

Introducción al reconocimiento del hablante 7 min

Ejercicio: Creación de una suscripción de Speaker Recognition API 10 min

Introducción a los conceptos de verificación del hablante 3 min

Administración de perfiles de verificación del hablante 7 min

Introducción a la identificación del hablante **5 min**

Administración de perfiles de identificación del hablante **7 min**

Implementación del reconocimiento del hablante **7 min**

Comprobación de conocimientos: reconocimiento del hablante
3 min

Resumen **2 min**

INTRODUCCIÓN

Las **Speaker Recognition** API de Azure Cognitive Services proporcionan algoritmos avanzados para la verificación e identificación del hablante. Estas API permiten incorporar autorización de voz en las aplicaciones e identificar voces específicas.

Nota

Este módulo requiere una suscripción de Azure. Los servicios que cree y use son gratuitos, pero necesitará una suscripción activa o una versión de prueba para completar los ejercicios. Si no tiene una suscripción a Azure, cree una [cuenta gratuita](#) antes de empezar.

Objetivos de este módulo:

- Conocer el concepto de reconocimiento del hablante.
- Obtener información sobre los conceptos relacionados con el reconocimiento del hablante.
- Explorar las Speaker Recognition API.

Supervisar los recursos

INTRODUCCIÓN AL RECONOCIMIENTO DEL HABLANTE

Las **Speaker Recognition** API de Azure Cognitive Services usan aprendizaje automático e inteligencia artificial para proporcionar servicios que:

- identifican a hablantes individuales;
- usan la voz para la autenticación.

Estos servicios se pueden integrar en aplicaciones y servicios mediante cualquier lenguaje de programación o sistema operativo.

5

El reconocimiento del hablante se divide en dos categorías:

- **Verificación del hablante:** los usuarios se verifican y autentican de forma automática a través de su voz.
- **Identificación del hablante:** la voz se compara con un grupo de posibles hablantes para identificar automáticamente a la persona que habla en un archivo de audio.

Speaker Recognition API

Todas las formas de vida parecen tener métodos de comunicación integrados y biológicamente adaptados. Con independencia de que

sea una ballena azul, un abejorro, una cigarra o un ser humano, cada "voz" es única. En nuestro planeta hay más 7,5 mil millones de voces humanas, cada una de ellas única y diferente al resto.

A través de simples llamadas de servicio basadas en REST, las Speaker Recognition API proporcionan algoritmos que le permiten:

- Verificar e identificar la voz humana.
- Organizar las voces en perfiles fáciles de administrar.

Los métodos proporcionados por las Speaker Recognition API se dividen en tres categorías:

1. **Verificación:** se comprueba la probabilidad de que dos voces pertenezcan a la misma persona.
2. **Identificación y reconocimiento:** se determina si una voz coincide con otra voz conocida.
3. **Inscripción:** se registran voces para verificarlas o identificarlas.

Las Speaker Recognition API permiten aprovechar la eficacia de la inteligencia artificial para:

- autenticar a los usuarios antes de permitirles acceder a aplicaciones y servicios;
- identificar hablantes en secuencias de vídeo;
- identificar hablantes en escenarios de chat en tiempo real;
- mejorar la seguridad de recursos confidenciales;

- reemplazar los complejos sistemas de autenticación heredados.

El reconocimiento del hablante a menudo se conoce como "reconocimiento de voz". El reconocimiento del hablante es el proceso de identificación de una persona en base a las características de voz de esa persona.

Nota

Hay una diferencia entre el *reconocimiento del hablante* y el *reconocimiento de voz*. El reconocimiento de voz intenta determinar el *contenido de las palabras habladas*, y el reconocimiento del hablante intenta identificar al *hablante del contenido*.

Acústica excelente

En el reconocimiento del hablante se evalúan las características acústicas vitales de voz que difieren entre individuos. Estos patrones acústicos vienen determinados por aspectos como el tamaño y la forma de la garganta y la boca. Cada individuo también tiene unos determinados patrones de voz, como el estilo de habla y el tono de voz.

Como resultado del progreso científico en el campo de verificación del hablante, ahora el reconocimiento del hablante se clasifica como una "biométrica del comportamiento". Como sucede con la diferencia entre el reconocimiento de voz y el del hablante, hay una diferencia entre el acto de *verificación* del hablante y el de *identificación*:

- **Verificación del hablante:** un intento por determinar si un hablante tiene una *identidad "afirmada" concreta*.
- **Identificación del hablante:** un intento por determinar la *identidad de un hablante desconocido*.

En ambos casos, es necesario disponer de voces o muestras de voces para poder realizar la comparación. El proceso de creación de muestras de voz se conoce como *inscripción*.

Inscripción

En el reconocimiento del hablante, la inscripción es el proceso de grabación previa de la voz del hablante para extraer los patrones acústicos y de voz que forman una *huella de voz*. La huella de voz a menudo se conoce como *plantilla de voz* o *modelo de voz*. En escenarios típicos de verificación del hablante, se graban previamente "expresiones" de voz de muestra que después se usan al comparar las huellas de voz.

El proceso de *identificación* del hablante es más difícil porque muchas veces una expresión se acaba de introducir por primera vez en el sistema. Por ejemplo, un sistema de reconocimiento del hablante podría entrenarse para identificar a un famoso presidente de Estados Unidos a partir de las muestras de los discursos de inauguración. Pero identificar a un presidente a partir de contenido nuevo, como un discurso sobre el estado de la nación, requiere la comparación del contenido hablado con varias huellas de voz a fin de determinar la mejor coincidencia.

Dado que los algoritmos en los que se basan la verificación y la identificación son diferentes, los procesos de inscripción para estos servicios también son independientes. En segundo plano, hay dos tipos de sistemas de reconocimiento del hablante: *dependiente de texto* e *independiente de texto*.

- **Dependiente de texto:** se usa para la inscripción en la verificación del hablante, donde los mensajes son conocidos y comunes entre todos los hablantes.
- **Independiente de texto:** se usa para la inscripción en la identificación del hablante, donde no hay ninguna cooperación (necesaria) por parte del hablante.

Nota

9

Para escenarios independientes de texto, la inscripción suele producirse sin conocimiento del usuario, ya que los sistemas de identificación del hablante no necesitan comparar lo que se dice en la inscripción.

El sistema dependiente de texto, el más controlado de todos, es fundamental para la verificación del hablante.

EJERCICIO: CREACIÓN DE UNA SUSCRIPCIÓN DE SPEAKER RECOGNITION API

Para poder usar Speaker Recognition API, tendrá que suscribirse al servicio y obtener una clave de suscripción.

Importante

Speaker Recognition API aún está en versión preliminar. Antes de que se finalice, espere cambios en algunas de las funciones.

10

1. Inicie sesión en [Azure Portal](#).
2. En el panel izquierdo, seleccione **+ Crear un recurso**.
3. En el cuadro **Buscar en Marketplace**, escriba **speaker recognition** (reconocimiento del hablante) y presione Entrar.
4. En los resultados de la búsqueda, seleccione **Speaker Recognition**.
5. Haga clic en **Crear**.
6. Escriba un nombre para la suscripción de Speaker Recognition API, como **SpeakerRecognition**.
7. Como **Ubicación**, elija la que le quede más cerca.

8. Como **Plan de tarifa**, seleccione un plan de tarifa.
9. Cree un grupo de recursos denominado **mslearn-speakerapi** para almacenar los recursos.
10. Revise y acepte el aviso de servicio y después haga clic en **Crear** para crear una suscripción a Speaker Recognition API.

Tras una breve pausa, la nueva suscripción a las Speaker Recognition API estará disponible y se generarán nuevas claves de suscripción para su uso en programación.

Nota

Los términos *clave de suscripción* y *clave de API* a menudo se usan como sinónimos en la documentación de Cognitive Services.

Para ver su nueva suscripción de Speaker Recognition, haga clic en **Grupos de recursos** en el panel izquierdo y, después, seleccione el grupo de recursos **mslearn speakerapi**.

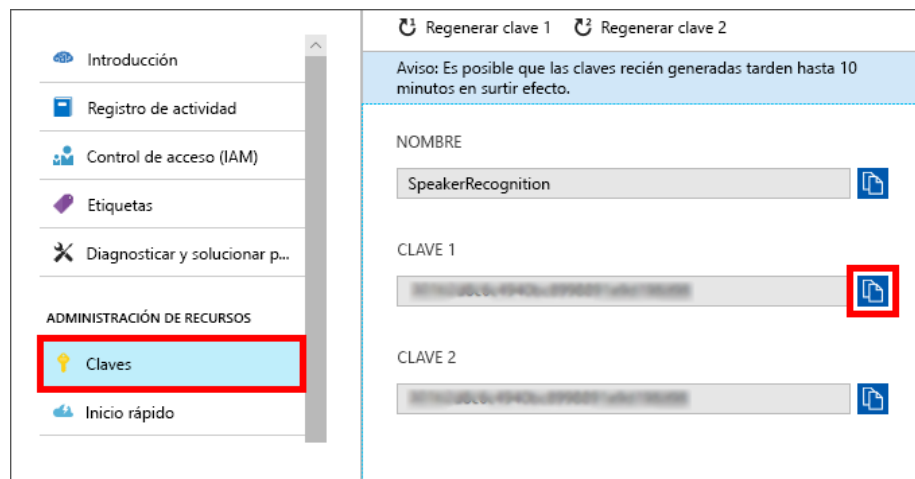
Para acceder a la suscripción de Speaker Recognition, necesita dos datos que puede obtener en Azure Portal:

- Una *clave de suscripción* que le permite acceder a su suscripción. Esta clave se debe pasar con cada llamada de servicio web que se realice a la API.
- El *punto de conexión* que expone el servicio en la red.

Copia de la clave de suscripción

Empecemos por obtener la clave de suscripción en Azure Portal:

1. En el panel izquierdo, haga clic en **Grupos de recursos** y, después, haga clic en el grupo de recursos que ha creado para Speaker Recognition API.
2. Escriba un nombre para la suscripción de Speaker Recognition (por ejemplo, **SpeakerRecognition**).
3. En **ADMINISTRACIÓN DE RECURSOS**, haga clic en **Claves** para ver las nuevas claves de suscripción.
4. Copie el valor de **CLAVE 1** en el portapapeles.




Sugerencia

Puede pegar este valor en un archivo de texto para usarlo más adelante en el curso.

Copia del punto de conexión de la suscripción

A continuación se obtendrá el punto de conexión de Azure Portal:

1. En el panel izquierdo, haga clic en **Introducción**.
2. Vaya a **Punto de conexión** y copie su valor en el portapapeles.

 Eliminar

Essentials ^

Grupo de recursos [\(cambiar\)](#)

[SpeechResources](#)

Estado

Activo

Ubicación

Oeste de EE. UU.

Nombre de suscripción [\(cambiar\)](#)

Tipo de API

Speaker Recognition API (vista previa)

Plan de tarifa

Gratis

Punto de conexión

<https://westus.api.cognitive.microsoft.com/spid/v1.0>

Administrar claves

[Mostrar claves de acceso...](#)

INTRODUCCIÓN A LOS CONCEPTOS DE VERIFICACIÓN DEL HABLANTE

Si es un veterano en viajes al extranjero, estará acostumbrado a presentar documentos emitidos por entidades gubernamentales cuando visita un país nuevo. La verificación del hablante es un proceso muy similar a entrar en un país nuevo con un pasaporte.

14



Cuando muestra su pasaporte al responsable de la oficina de inmigración, usted afirma ser la persona que aparece en la foto del pasaporte. Al igual que existen normativas de cómo se debe hacer esta foto, el proceso de verificación de voz normalmente tiene unas directrices sobre cuál es la "apariencia" de su huella de voz. En otras palabras, lo que dice para generar su huella de voz está regulado. La mayoría de los sistemas usan el concepto de *frases de verificación*.

Frases de verificación

El servicio Speaker Recognition usa frases de verificación predefinidas para determinar las características y los patrones acústicos y de voz pertinentes del hablante. En la actualidad, el servicio Speaker Recognition admite las 10 frases de verificación siguientes, para la configuración regional de inglés (Estados Unidos):

- "I am going to make them an offer they cannot refuse." (Voy a hacerles una oferta que no podrán rechazar)
- "Houston, we have a problem." (Houston, tenemos un problema)
- "My voice is my passport. Verify me." (Mi voz es mi pasaporte. Compruébeme)
- "Apple juice tastes funny after toothpaste." (El zumo de manzana sabe raro después de lavarse los dientes)
- "You can get in without your password." (Puede entrar sin su contraseña)
- "You can activate the security system now." (Ya puede activar el sistema de seguridad)
- "My voice is stronger than passwords." (Mi voz es más segura que las contraseñas)
- "My password is not your business." (Mi contraseña no es asunto suyo)
- "My name is unknown to you." (No sabe mi nombre)

- "Be yourself. Everyone else is already taken." (Sea usted mismo, todos los demás ya están en uso)

Nota

Speaker Recognition API no admite actualmente frases de verificación personalizadas.

La inscripción de verificación del hablante requiere la grabación y el envío de tres ejemplos de una frase idéntica. Por ejemplo, se le pide que repita la frase "Houston, tenemos un problema" tres veces y, después, que envíe esos ejemplos al sistema.

Después de enviar las frases de ejemplo idénticas, la inscripción se considera completada y ya puede verificar su identidad con esa frase.

Quizás se pregunte por qué se necesita una frase de verificación idéntica. ¿Por qué no puede el sistema analizar y evaluar los patrones acústicos y de voz del hablante para determinar si existe una coincidencia de voz? En realidad, el sistema puede identificar una coincidencia de voz sin necesidad de una frase de verificación idéntica.

Pero si se permitiese cualquier muestra de voz, el sistema ya no sería seguro. Alguien podría grabar su voz y usar esa grabación en el sistema. Pedirle que recuerde y use un ejemplo exacto de frase es muy similar a tener que recordar una contraseña, el nombre de soltera de su madre, el nombre de su primera mascota o la ciudad donde nació, pero mejor.

Una vez completada la inscripción, el sistema Speaker Recognition contiene el *perfil de verificación*.

Perfiles de verificación

TABLA 1

Propiedad	Valor
verificationProfileId	Un identificador de perfil único
locale	La región o la configuración regional de referencia cultural, como "en-US"
enrollmentsCount	El número total de inscripciones creadas
remainingEnrollmentsCount	El número restante de inscripciones necesarias para la finalización
createdDateTime	La fecha y la hora en que se ha creado el perfil
lastActionDateTime	La fecha y hora de la última acción que se ha realizado
enrollmentStatus	El estado de la inscripción

Un perfil de verificación es una agrupación de inscripciones dependientes de texto que se usa para la verificación del hablante. Dado que un hablante debe elegir una frase específica para la inscripción y la verificación, el perfil resultante contiene todo lo necesario para la verificación futura del hablante.

Sugerencia

TABLA 2

Valor	Estado de verificación
enrolling	Se está inscribiendo un perfil que aún <i>no está</i> listo para realizar la verificación.
training	Se está entrenando un perfil que aún <i>no está</i> listo para realizar la verificación.
enrolled	Se ha inscrito un perfil y <i>está</i> listo para realizar la verificación.

Aunque no sea absolutamente necesario, se recomienda usar el mismo micrófono para la inscripción y la verificación.

Las propiedades de un perfil de comprobación

En Speaker Recognition API, un perfil de verificación incluye las propiedades siguientes:

Propiedad de estado de inscripción

Aparte de `verificationProfileId`, la propiedad más importante de un perfil de verificación es `enrollmentStatus`. Este valor muestra el estado actual de la inscripción y la preparación para la verificación del hablante:

ADMINISTRACIÓN DE PERFILES DE VERIFICACIÓN DEL HABLANTE

La creación de perfiles de verificación es un proceso sencillo que utiliza los métodos de administración de verificación de Speaker Verification API.

Uso de Speaker Verification API para administrar perfiles de verificación

La administración de perfiles de verificación del hablante con Speaker Verification API es un proceso de dos fases:

- Enviar una solicitud web autorizada al punto de conexión de la suscripción.
- Observar la información que se devuelve tras la llamada.

Al igual que la mayoría de los servicios del conjunto de aplicaciones de Azure Cognitive Services, Speaker Verification API devuelve información en formato JSON estándar.

Los métodos principales que se usan para crear y administrar perfiles de verificación son los métodos Create Profile y Create Enrollment de Speaker Verification API. El método Create Profile acepta un parámetro

de región o configuración regional para crear un perfil basado en una configuración regional específica, y devuelve un identificador único para el perfil. Pero el método Create Enrollment es donde se produce un trabajo considerable.

El método Create Enrollment acepta un archivo de audio (como una carga de archivo binario) que *debe* cumplir los criterios siguientes:

- Contenedor: WAV
- Codificación: PCM
- Velocidad: 16 K
- Formato de muestreo: 16 bits
- Canales: Mono

El método Create Enrollment requiere pasar un parámetro único al servicio. Este parámetro `verificationProfileId`, que se devuelve desde el método Create Profile, identifica a qué perfil debe aplicarse el envío de la inscripción.

Normalmente, en la cadena de consulta se envía un parámetro pasado al método Create Enrollment, como parte de la dirección URL del punto de conexión:

textCopiar

<https://westus.api.cognitive.microsoft.com/spid/v1.0/verificationProfiles/{verificationProfileId}/enroll>

- Sugerencia

Cuando se proporciona el audio de inscripción a través de una solicitud basada en REST, el archivo de audio debe tener al menos 1 segundo de duración y no más de 15 segundos.

Envío de una carga de audio

Para enviar una carga de audio al método Create Enrollment, la solicitud web debe incluir un archivo binario, como una secuencia o matriz de bytes. Después debe enviarse la solicitud a través de un método HTTP POST estándar.

Carga del archivo binario

Para crear y enviar una carga de archivo binario al método Create Enrollment, se usan métodos estándar específicos del lenguaje para crear y enviar contenido binario. Por ejemplo, en C#, una carga binaria podría proceder de un archivo de audio (que contenga una grabación de voz de la frase de verificación) ubicado en un equipo local:

C#Copiar

```
string uri =  
"https://westus.api.cognitive.microsoft.com/spid/v1.0/verificationProfiles/{verificationProfileId}/enroll";
```

```
FileStream fileStream = new FileStream(@"c:\\Sample  
Audio\\enrollment_01.wav", FileMode.Open, FileAccess.Read);
```

```
BinaryReader binaryReader = new BinaryReader(fileStream);
```

```
var bytes = binaryReader.ReadBytes((int)fileStream.Length);
```

```

using (ByteArrayContent content = new ByteArrayContent(bytes))

{

    content.Headers.ContentType = new
    MediaTypeHeaderValue("application/octet-stream");

    //POST the binary content to the endpoint

    response = await client.PostAsync(uri, content);

    string contentString = await response.Content.ReadAsStringAsync();

}

```

El tipo de contenido application/octet-stream debe estar incluido en el encabezado de solicitud para que el método sepa qué tipo de contenido se va a suministrar.

No es necesario usar un parámetro para proporcionar la frase de verificación seleccionada. Speaker Verification API extrae la frase del archivo de audio y presenta este valor como un modo de proteger la verificación.

Valores devueltos

Los resultados devueltos por Speaker Verification API se incluyen en cargas de JSON. Este es un ejemplo de la información devuelta por el método Create Enrollment:

JSONCopiar

```
{  
  
  "enrollmentStatus" : "Enrolled",  
  
  "enrollmentsCount":3,  
  
  "remainingEnrollments" : 0,  
  
  "phrase" : "<Recognized verification phrase>"  
  
}
```

- Nota

El valor "<Recognized verification phrase>" se rellena con la frase de verificación extraída.

La administración de perfiles de verificación a través de Speaker Verification API es sencilla, tanto como la administración de perfiles de identificación.

INTRODUCCIÓN A LA IDENTIFICACIÓN DEL HABLANTE

Si el uso de la verificación del hablante es similar al proceso de mostrar un pasaporte, la identificación del hablante es como la comparación de huellas dactilares, que pregunta lo siguiente: "De todas las huellas digitales conocidas por nuestro sistema, ¿cuál coincide con esta huella digital?"



La identificación del hablante es la capacidad para determinar la persona que habla en un audio según los patrones acústicos y de voz, como el estilo de habla y el tono de voz.

En un contexto de orden público, un sistema de identificación del hablante puede emplear una huella de voz *independiente de texto*, de forma similar a como se utilizan las huellas digitales. De este modo, la identificación del hablante es básicamente "abierta". En el ejemplo de

las huellas, tomar la huella dactilar en cualquier momento introduce esa huella en la "inscripción".

La identificación del hablante tiene muchas otras aplicaciones fuera del ámbito de la aplicación de la ley y el análisis forense. Como sucede en escenarios de verificación del hablante, cuando se completa la inscripción, el servicio Speaker Identification contiene el *perfil de identificación* del hablante.

Perfiles de identificación

Un perfil de verificación es una agrupación de inscripciones independientes de texto que se usan para la identificación del hablante. Al contrario de lo que sucede en la verificación del hablante, un hablante no tiene que elegir una frase concreta para usarla en la inscripción o la identificación.

Nota

A diferencia de la verificación del hablante, no es razonable crear el audio de identificación del hablante con el mismo micrófono (o incluso crearlo en la misma década), por lo que no hay requisitos ni recomendaciones sobre las especificaciones de grabación o el dispositivo de entrada. Pero el archivo real enviado debe cumplir los requisitos en cuanto a formato y velocidad de bits.

El perfil

En Speaker Identification API, un perfil de identificación incluye las propiedades siguientes:

TABLA 1

Propiedad	Valor
identificationProfileId	Un identificador de perfil único
locale	La región o la configuración regional de referencia cultural, como "en-US"
enrollmentSpeechTime	El número total de segundos de voz "útil" detectados en todos los archivos de audio de inscripción proporcionados
remainingEnrollmentSpeechTime	El número restante de segundos de voz necesarios para una inscripción correcta
createdDateTime	La fecha y la hora en que se ha creado el perfil
lastActionDateTime	La fecha y hora de la última acción que se ha realizado
enrollmentStatus	El estado de la inscripción

Estado de inscripción

Aparte de identificationProfileId, la propiedad más importante de un perfil de identificación es enrollmentStatus. Este valor muestra el estado actual de la inscripción y la preparación para la identificación del hablante:

TABLA 2

Valor	Estado de inscripción
enrolling	Un perfil se está inscribiendo actualmente y <i>no está</i> preparado para la identificación.
training	Un perfil se está entrenando actualmente y <i>no está</i> preparado para la identificación.
enrolled	El perfil está inscrito y <i>preparado</i> para la identificación.

ADMINISTRACIÓN DE PERFILES DE IDENTIFICACIÓN DEL HABLANTE

La **creación de** perfiles de identificación es un proceso sencillo que utiliza los métodos de administración de identificación de Speaker Identification API.

Uso de Speaker Identification API para administrar perfiles de identificación

28

Al igual que con los perfiles de verificación, para administrar perfiles de identificación del hablante con Speaker Identification API hay que hacer lo siguiente:

- Enviar una solicitud web autorizada al punto de conexión de la suscripción.
- Observar la información que se devuelve tras la llamada.

Los métodos principales que se usan para crear y administrar perfiles de identificación son los métodos Create Profile y Create Enrollment de Speaker Identification API. Como en la verificación del hablante, el método Create Profile acepta un parámetro de región o configuración regional para crear un perfil basado en una configuración regional específica, y devuelve un identificador único

para el perfil. Pero el método Create Enrollment es donde se produce un trabajo considerable.

El método Create Enrollment acepta un archivo de audio (como una carga de archivo binario) que *debe* cumplir los mismos criterios que los archivos usados para la verificación del hablante:

TABLA 1

Propiedad	Valor requerido
Contenedor	WAV
Codificación	PCM
Velocidad	16 K
Formato de muestreo	16 bits
Canales	Mono

El método Create Enrollment requiere pasar un parámetro único al servicio. Este parámetro `identificationProfileId`, devuelto por el método Create Profile, identifica a qué perfil debe aplicarse el envío de la inscripción.

También se puede pasar un parámetro opcional `shortAudio` para indicar al servicio que omita los requisitos de duración del audio. Este parámetro es un valor de "sí" o "no" (booleano) que indica al servicio que renuncie al límite de audio mínimo recomendado necesario para la inscripción.

Normalmente, en la cadena de consulta se envía un parámetro pasado al método Create Enrollment de Speaker Identification API, como parte de la dirección URL del punto de conexión. En este ejemplo se incluye el parámetro opcional shortAudio:

textCopiar

```
https://westus.api.cognitive.microsoft.com/spid/v1.0/identificationProfiles/{identificationProfileId}/enroll?shortAudio=true
```

- Nota

Si se inscribe un hablante para su identificación sin usar el parámetro shortAudio, los archivos de audio deben tener una duración mínima de 5 segundos y máxima de 5 minutos. La cantidad *mínima* recomendada de voz acumulada para la inscripción, después de quitar los silencios, es de 30 segundos. Después de acumular 30 segundos de voz, el estado de inscripción del perfil cambia de enrolling a enrolled para indicar que está listo para la identificación. Si se usa el parámetro shortAudio, los archivos de audio pueden tener hasta 1 segundo de duración, pero la precisión de la identificación del hablante puede verse afectada.

Envío de una carga de audio

Para enviar una carga de audio al método Create Enrollment de Speaker Identification, es necesario incluir en la solicitud web un archivo binario (como una secuencia o matriz de bytes) y, después, enviar la solicitud a través de un método HTTP POST estándar.

Carga del archivo binario

Para crear y enviar una carga de archivo binario al método Create Enrollment, se usan métodos estándar específicos del lenguaje para crear y enviar contenido binario. Por ejemplo, en C#, una carga binaria podría proceder de un archivo de audio (que contenga una grabación de voz de la persona que habla) ubicado en un equipo local:

C#Copiar

```
string uri =
"https://westus.api.cognitive.microsoft.com/spid/v1.0/identificationPr
ofiles/{identificationProfileId}/enroll?shortAudio=true";

FileStream fileStream = new FileStream(@"c:\\Sample
Audio\\enrollment_01.wav", FileMode.Open, FileAccess.Read);

BinaryReader binaryReader = new BinaryReader(fileStream);

var bytes = binaryReader.ReadBytes((int)fileStream.Length);

using (ByteArrayContent content = new ByteArrayContent(bytes))
{
    content.Headers.ContentType = new
MediaTypeHeaderValue("application/octet-stream");

    //POST the binary content to the endpoint

    response = await client.PostAsync(uri, content);
```

```
string contentString = await response.Content.ReadAsStringAsync();  
  
}
```

En el ejemplo anterior, el parámetro opcional `shortAudio` indica al método que renuncie al límite recomendado de audio mínimo necesario para la inscripción y que permita que los archivos de audio duren tan solo un segundo.

Valores devueltos

Si tiene archivos de audio de larga duración, crear las inscripciones de identificación del hablante puede ser un proceso lento. Dado que la inscripción es un proceso potencialmente lento, el método `Create Enrollment` es un método de resultado y ubicación de la operación.

Para obtener el estado de una operación de inscripción, es preciso:

- Enviar una solicitud inicial, con la carga especificada.
- Obtener el valor de **ubicación de la operación** de la solicitud inicial.
- Sondear el valor de **ubicación de la operación** que se devuelve, hasta que el valor de **resultado de la operación** indique que la operación se ha realizado correctamente.
- Llamar al método `Get Profile` en la `Speaker Identification API`.

Cuando el proceso de inscripción se realiza correctamente, los resultados devueltos por el método `Get Profile` están en un objeto

JSON con formato correcto o en una matriz que tiene el aspecto del ejemplo siguiente:

JSONCopiar

```
{  
  
  "identificationProfileId": "e30148ba-9078-4b88-a53c-  
d47888ba64ce",  
  
  "locale": "en-us",  
  
  "enrollmentSpeechTime": 3.38,  
  
  "remainingEnrollmentSpeechTime": 0.0,  
  
  "createdDateTime": "2018-02-18T00:39:58.567Z",  
  
  "lastActionDateTime": "2018-02-18T00:40:00.614Z",  
  
  "enrollmentStatus": "Enrolled"  
}
```

Una vez aplicado un perfil de verificación o identificación del hablante, ya se puede usar Azure Cognitive Services para reconocer al hablante. Para facilitar las pruebas de esta característica, compruebe que sigue las instrucciones para crear la cuenta de Cognitive Services y, después, use la clave de servicio y visite la [página de Speaker Recognition API](#), donde encontrará más información y podrá seleccionar el botón Abrir consola de pruebas de API cuando haya preparado la carga de su propia muestra de audio y la prueba de las características de identificación. La aplicación de pruebas de API permite realizar pruebas

sin tener que disponer de una aplicación de un extremo a otro completa.

IMPLEMENTACIÓN DEL RECONOCIMIENTO DEL HABLANTE

El **proceso** de verificación del hablante consta de tres pasos principales:

1. Crear un perfil de verificación
2. Crear tres inscripciones con una frase de verificación
3. Intentar verificar a un hablante

Para verificar un hablante, se debe enviar una solicitud web autorizada al punto de conexión de la suscripción y observar la información devuelta por la llamada a través del método Verify de Speaker Verification API.

TABLA 1

Propiedad	Valor requerido
Contenedor	WAV
Codificación	PCM

Velocidad	16 K
Formato de muestreo	16 bits
Canales	Mono

El método Verify acepta un archivo de audio (como una carga de archivo binario) que *debe* cumplir los mismos criterios que los archivos usados para la inscripción de la verificación y la identificación del hablante. Repasemos estos requisitos de archivo de audio:

El método Verify no acepta ningún parámetro adicional. Solo requiere un clip de audio (como un archivo binario) para la verificación. En C#, una carga binaria puede proceder de un archivo de audio (que contenga una grabación de voz "nueva" de la frase de verificación), ubicado en un equipo local:

C#Copiar

```
string uri =
"https://westus.api.cognitive.microsoft.com/spid/v1.0/verify?verificationProfileId={verificationProfileId}";
```

```
FileStream fileStream = new FileStream(@"c:\\Sample
Audio\\verification_01.wav", FileMode.Open, FileAccess.Read);
```

```
BinaryReader binaryReader = new BinaryReader(fileStream);
```

```
var bytes = binaryReader.ReadBytes((int)fileStream.Length);
```

```

using (ByteArrayContent content = new ByteArrayContent(bytes))

{

    content.Headers.ContentType = new
    MediaTypeHeaderValue("application/octet-stream");

    //POST the binary content to the endpoint

    response = await client.PostAsync(uri, content);

    string contentString = await response.Content.ReadAsStringAsync();

}

```

Valores devueltos

La información devuelta desde una llamada al método Verify proporciona toda la información necesaria para verificar a un hablante:

- Result: un valor de Accept o Reject.
- Confidence: un valor de Low, Normal o High.
- Phrase: la frase de verificación reconocida del archivo de audio

Una vez realizada la verificación, se devuelve una matriz o un objeto JSON con el formato correcto:

JSONCopiar

```

{

    "result" : "Accept",

```

```
"confidence" : "Normal",  
  
"phrase": "<Recognized phrase>"  
  
}
```

Nota

El valor <Recognized phrase> se rellena con la frase extraída que se reconoció en el audio.

Según el escenario, la información devuelta podría requerir lógica adicional en su aplicación o servicio. Por ejemplo, en entornos altamente confidenciales, un nivel de confianza Normal podría indicar la necesidad de una nueva verificación hasta que se notifique un nivel de confianza High. En caso contrario, un valor Accept podría ser suficiente por sí mismo.

Una vez que esté familiarizado con el proceso de verificación del hablante, verá que el proceso de identificación del hablante es similar.

Identificación del hablante

Al igual que en la verificación del hablante, el proceso de identificación del hablante consta de tres pasos principales:

1. Crear un perfil de identificación
2. Crear una inscripción de identificación
3. Intentar identificar a un hablante

Al igual que en la verificación del hablante, para identificar un hablante se debe enviar una solicitud web autorizada al punto de conexión de

la suscripción y observar la información devuelta por la llamada a través del método Identify de Speaker Identification API. El archivo de audio que envíe en su solicitud debe cumplir los mismos criterios que todos los demás archivos de audio del sistema.

El método Identify tampoco toma parámetros adicionales, solo un clip de audio (como un archivo binario) que contiene cualquier forma de contenido hablado. En C#, una carga binaria podría proceder de un archivo de audio (que contenga una grabación de voz "nueva"), ubicado en un equipo local:

C#Copiar

```
string uri =
"https://westus.api.cognitive.microsoft.com/spid/v1.0/identify?identificationProfileId={identificationProfileId}";
```

```
FileStream fileStream = new FileStream(@"c:\Sample
Audio\identification_01.wav", FileMode.Open, FileAccess.Read);
```

```
BinaryReader binaryReader = new BinaryReader(fileStream);
```

```
var bytes = binaryReader.ReadBytes((int)fileStream.Length);
```

```
using (ByteArrayContent content = new ByteArrayContent(bytes))
```

```
{
```

```
    content.Headers.ContentType = new
    MediaTypeHeaderValue("application/octet-stream");
```

```
//POST the binary content to the endpoint.

response = await client.PostAsync(uri, content);

string contentString = await response.Content.ReadAsStringAsync();

}
```

Valores devueltos de identificación

La información devuelta por una llamada al método Identify proporciona toda la información necesaria para identificar a un hablante:

- identifiedProfileId: el valor de identificationProfileId del perfil del hablante identificado.
- Confidence: un valor de Low, Normal o High.

Cuando ha finalizado la identificación, se devuelve un objeto JSON:

JSONCopiar

```
{

  "identifiedProfileId" : "00000000-0000-0000-0000-000000000000",

  "confidence" : "Normal"

}
```


Importante

Si la propiedad `identifiedProfileId` devuelve un valor de 00000000-0000-0000-0000-000000000000, significa que no se ha identificado ningún perfil de identificación del hablante. En otras palabras, el archivo de audio que debe identificarse no coincide con ninguno de los perfiles proporcionados para la identificación del hablante.