



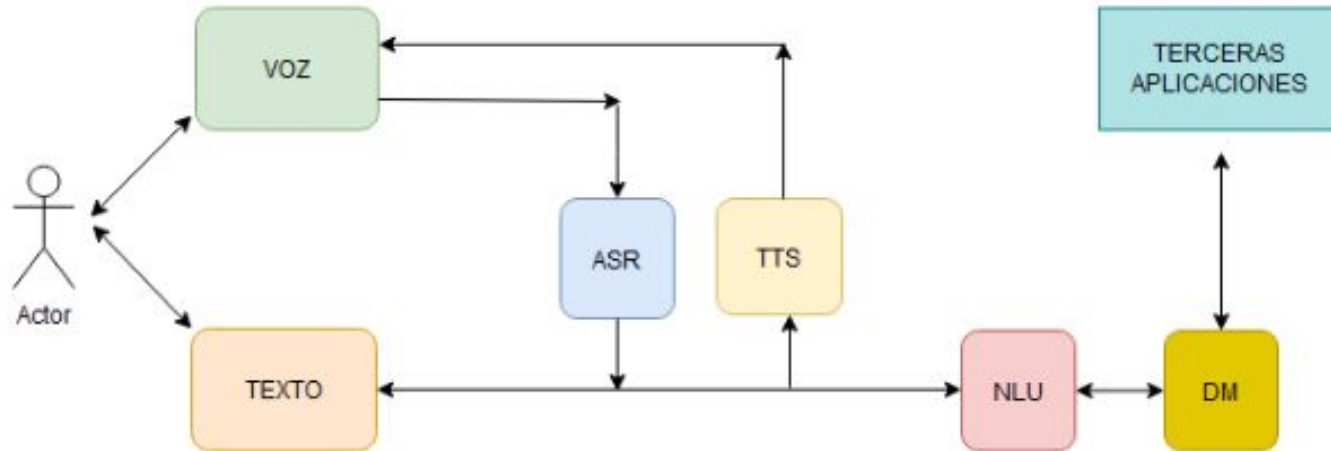
UT.5 PROCESAMIENTO DEL LENGUAJE NATURAL: TEXTO A VOZ

Bloques de la unidad:



1. Text to Speech (TTS).
2. Speech Synthesis Markup Language (SSML).
3. Laboratorio 1: SSML.

1. Text to speech (TTS)



1. Text to speech (TTS)



Ejemplo de un proyecto de IA empresarial:

<https://reunir.unir.net/handle/123456789/13772>

1. Text to speech (TTS)



La conversión de texto a voz usa redes neuronales profundas para hacer que las voces de los equipos resulten prácticamente imposibles de distinguir de las grabaciones de las personas. Gracias a la clara articulación de las palabras, la funcionalidad de texto a voz neuronal reduce significativamente la fatiga auditiva cuando los usuarios interactúan con sistemas de inteligencia artificial.

1. Text to speech (TTS)



Los casos de uso más comunes son:

- Las aplicaciones móviles.
- Lectores de noticias, juegos.
- Plataformas de recursos de aprendizaje electrónicos.
- Aplicaciones de accesibilidad para personas con discapacidad.
- Segmento de Internet de las cosas (IoT), en rápido crecimiento.

1. Text to speech (TTS)



Las características principales son:

- Síntesis de voz en tiempo real.
- Síntesis asincrónica de audio de larga duración.
- Voces neuronales pregeneradas.
- Ajuste de la salida de texto a voz con SSML.
- Visemas.

1. Text to speech (TTS)



Los visemas son los principales planteamientos de la voz observada, incluida la posición de los labios, mandíbula y lengua al producir un fonema determinado. Los visemas tienen una correlación fuerte con las voces y fonemas.

Mediante el uso de eventos de visema en el SDK de Voz, puede generar datos de animación facial. Estos datos se pueden usar para animar caras en la comunicación con lectura de labios, la educación, el entretenimiento y el servicio de atención al cliente. (Ojo novedoso y normalmente sólo en inglés)

1. Text to speech (TTS)



2. Speech Synthesis Markup Language (SSML).



El lenguaje de marcado de síntesis de voz (SSML) es un lenguaje de marcado basado en XML que puede utilizar para ajustar los atributos de salida de texto a voz, como el tono, la pronunciación, la velocidad de habla, el volumen, etc. Le ofrece más control y flexibilidad que la introducción de texto sin formato.

2. Speech Synthesis Markup Language (SSML).

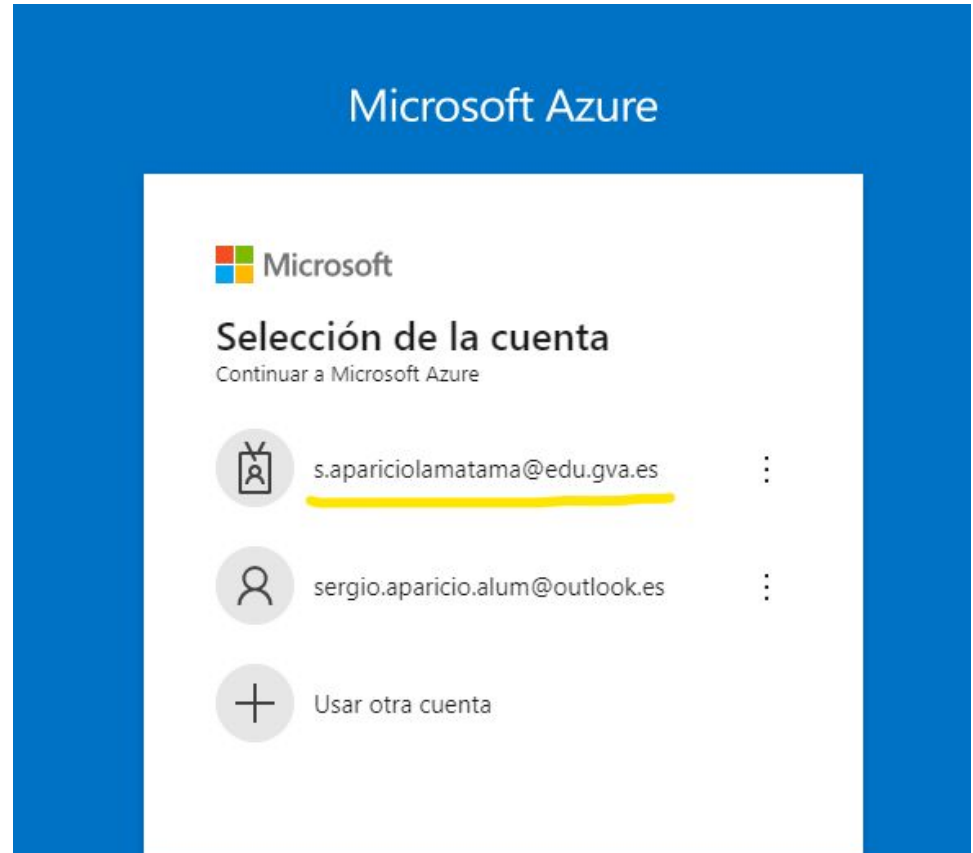


Nosotros vamos a utilizar Azure y nos permite utilizar SSML de varias maneras:

- La herramienta de Creación de contenido de audio le permite crear texto sin formato y SSML en Speech Studio. La API de síntesis por lotes acepta SSML mediante la propiedad inputs.
- La CLI para Voz acepta SSML a través del argumento de línea de comandos `spx synthesize --ssml SSML`.
- El SDK para Voz acepta SSML a través del método "speak" SSML en los diferentes idiomas soportados.


2. Speech Synthesis Markup Language (SSML).

Vamos a utilizar Speech Studio, lo primero que tenemos que hacer es conectarnos a Azure con la cuenta de la gva.











2. Speech Synthesis Markup Language (SSML).


Buscamos “servicio de voz”

Azure AI services | Servicio de Voz  ...










Azure AI services



Buscar

+ Crear  Administrar planes de compromiso  Administrar recursos eliminados  Administrar vista  Actualizar  Exportar a CSV  Abrir consulta  Asignar etiquetas  Eliminar

Filtrar por cualquier ca... Suscripción es igual a **todo** Tipo es igual a **todo** Grupo de recursos es igual a **todo** Ubicación es igual a **todo**  Agregar filtro

Mostrando de 0 a 0 de 0 registros.

Nombre 	Variante 	Ubicación 	Nombre de dominio per... 	Plan de tarifa 	Estado 	Fecha de creación 
<div><p>No hay Servicios de voz para mostrar</p><p>Intente cambiar o borrar los filtros.</p><p>Crear Servicio de Voz</p><p>Más información </p></div>						

Sin agrupar   Vista de lista

Azure AI services

- Información general
- All Azure AI services
- Azure AI services
- Azure OpenAI
- AI Search
- Computer Vision
- API de Face
- Custom Vision
- Servicio de Voz**
- Servicio de Lenguaje
- Traductor
- Document intelligence
- Servicios de bot
- Anomaly Detector
- Content Moderator
- Análisis prospectivo de estado

2. Speech Synthesis Markup Language (SSML).

Crear Servicios de voz

Basics Red Identity Etiquetas Revisar y crear

Transcriba la voz en texto legible en el que se puedan realizar búsquedas. Agregue traducciones de voz en tiempo real a sus aplicaciones y servicios. Convierta texto en audio casi en tiempo real. Cree rápidamente aplicaciones y servicios habilitados para voz mediante los lenguajes de programación con los que ya trabaja. Personalice los sistemas de voz para optimizar la calidad en escenarios específicos.

[Más información](#)

Detalles del proyecto

Suscripción * ⓘ

Grupo de recursos * ⓘ

[Crear nuevo](#)

Detalles de la instancia

Región ⓘ

Nombre * ⓘ

Plan de tarifa * ⓘ

❌ El valor no debe estar vacío.

[Ver todos los detalles de los precios](#)

2. Speech Synthesis Markup Language (SSML).

Basics Red Identity Etiquetas Revisar y crear

Transcriba la voz en texto legible en el que se puedan realizar búsquedas. Agregue traducciones de voz en tiempo real a sus aplicaciones y servicios. Convierta texto en audio casi en tiempo real. Cree rápidamente aplicaciones y servicios habilitados para voz mediante los lenguajes de programación con los que ya trabaja. Personalice los sistemas de voz para optimizar la calidad en escenarios específicos.

[Más información](#)

Detalles del proyecto

Suscripción * ⓘ

Azure for Students ▼

Grupo de recursos * ⓘ

(Nuevo) PruebaModelos ▼

[Crear nuevo](#)

Detalles de la instancia

Región ⓘ

East US ▼

Nombre * ⓘ

Estenombretienezserunico ✓

Plan de tarifa * ⓘ

Free F0 ▼

[Ver todos los detalles de los precios](#)

2. Speech Synthesis Markup Language (SSML).

Basics

Red

Identity

Etiquetas

Revisar y crear



Configure network security for your Azure AI services resource.



Tipo *


- ☒ Todas las redes, incluso Internet, pueden acceder a este recurso.
- ☐ Selected networks, configure network security for your Azure AI services resource.
- ☐ Se ha deshabilitado, por lo que ninguna red puede acceder a este recurso. La única manera de hacerlo es configurar conexiones de puntos de conexión privados.

2. Speech Synthesis Markup Language (SSML).

Basics Red Identity Etiquetas Revisar y crear

Identidad administrada asignada por el sistema

Permite habilitar la identidad asignada por el sistema para conceder acceso al recurso a otros recursos existentes.



Estado 

☒ Desactivado

☐ Activado

Identidad administrada asignada por el usuario

Permite agregar identidades asignadas por el usuario para conceder acceso al recurso a otros recursos existentes.

 Agregar  Quitar

Nombre	↑↓ grupo de recursos	↑↓ suscripción	↑↓
No hay identidades administradas asignadas por el usuario asignadas a este recurso. Seleccione "Agregar" para agre...			

2. Speech Synthesis Markup Language (SSML).

Basics Red Identity Etiquetas Revisar y crear

Las etiquetas son pares nombre-valor que permiten categorizar los recursos y ver una facturación consolidada mediante la aplicación de la misma etiqueta en varios recursos y grupos de recursos. [Más información sobre las etiquetas](#)

Tenga en cuenta que si crea etiquetas y, después, cambia la configuración de los recursos en otras pestañas, las etiquetas se actualizan automáticamente.

Nombre ⓘ

Valor ⓘ

Recurso

:

Azure AI services

2. Speech Synthesis Markup Language (SSML).

Basics Red Identity Etiquetas Revisar y crear

📄 [Descargar una plantilla para la automatización](#)

TÉRMINOS

Al hacer clic en "Crear", (a) acepto los términos legales y las declaraciones de privacidad relacionados con cada oferta de Marketplace que se enumeró previamente; (b) autorizo a Microsoft a facturar con mi método de pago actual las cuotas relacionadas con las ofertas, con la misma frecuencia de facturación que mi suscripción de Azure; y (c) autorizo a Microsoft a compartir mi información de contacto y los datos de transacción y uso con los proveedores de dichas ofertas. Microsoft no proporciona derechos sobre ofertas de terceros. Para obtener información adicional, consulte [los Términos de Azure Marketplace](#).

Basics

Suscripción	Azure for Students
Grupo de recursos	PruebaModelos
Región	East US
Nombre	Estenombretienequeserunico
Plan de tarifa	Free F0

Red

Tipo	Todas las redes, incluso Internet, pueden acceder a este recurso.
------	---

Identity

Tipo de identidad	None
-------------------	------

< Anterior

Siguiente

Crear

2. Speech Synthesis Markup Language (SSML).

 Eliminar  Cancelar  Volver a implementar  Descargar  Actualizar

✓ Se completó la implementación



Nombre de implementación : Microsoft.CognitiveServicesSpeechServices-20240211192615

Suscripción : [Azure for Students](#)

Grupo de recursos : [PruebaModelos](#)

> Detalles de implementación

✓ Pasos siguientes

[Ir al recurso](#)

Enviar comentarios



Cuéntenos su experiencia con la implementación

2. Speech Synthesis Markup Language (SSML).

Empezar

Comience a usar el recurso en Speech Studio



Pruebe todos los casos de uso y vea otras herramientas personalizadas para crear modelos de IA del habla


[Ir a Speech Studio](#)

Claves y punto de conexión

i These keys are used to access your Azure AI services API. Do not share your keys. Store them securely– for example, using Azure Key Vault. We also recommend regenerating these keys regularly. Only one key is necessary to make an API call. When regenerating the first key, you can use the second key for continued access to the service.


Mostrar claves

Clave 1

..... 


Clave 2

..... 

Ubicación o región 

eastus 

Extremo

https://eastus.api.cognitive.microsoft.com/ 

2. Speech Synthesis Markup Language (SSML).



¿ TODOS YA
ESTAMOS EN ESTE
PUNTO ?

2. Speech Synthesis Markup Language (SSML).

Vamos a Texto a Voz

Texto a voz

Compile aplicaciones y servicios que hablen de forma natural con más de 400 voces en 140 idiomas y dialectos. Cree una voz personalizada para diferenciar su marca y utilice varios estilos de habla para dar un sentido de emoción a su contenido hablado. [Más información sobre texto a voz](#)



Galería de voz

Busque voces expresivas con un habla similar a la humana para encontrar el orador perfecto para su proyecto.

[Probar la Galería de voz](#)



Voz personalizada

Use sus propias grabaciones de audio para crear una voz distinta y única para sus aplicaciones de texto a voz.


[Iniciar un proyecto de habla personalizada](#)



Voz personal *Vista previa*

Cree fácilmente una voz de inteligencia artificial a partir de una muestra de voz humana limpia, lo que proporciona a los usuarios una experiencia de voz personalizada en 100 idiomas.

[Probar Personal Voice](#)

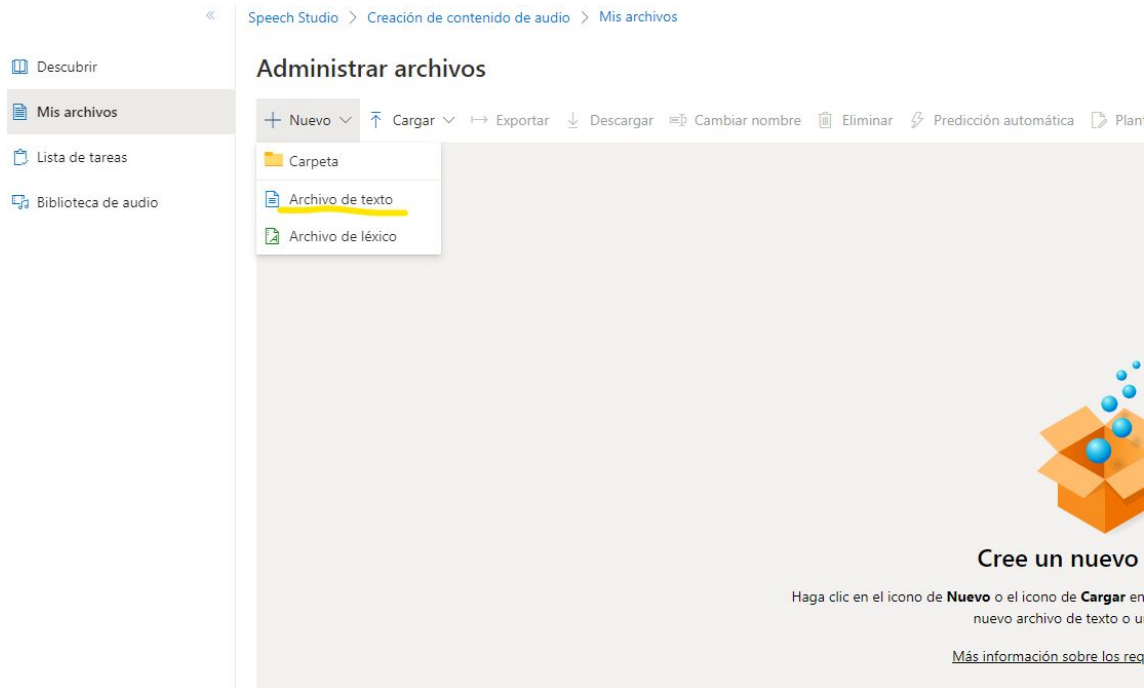


Creación de contenido de audio

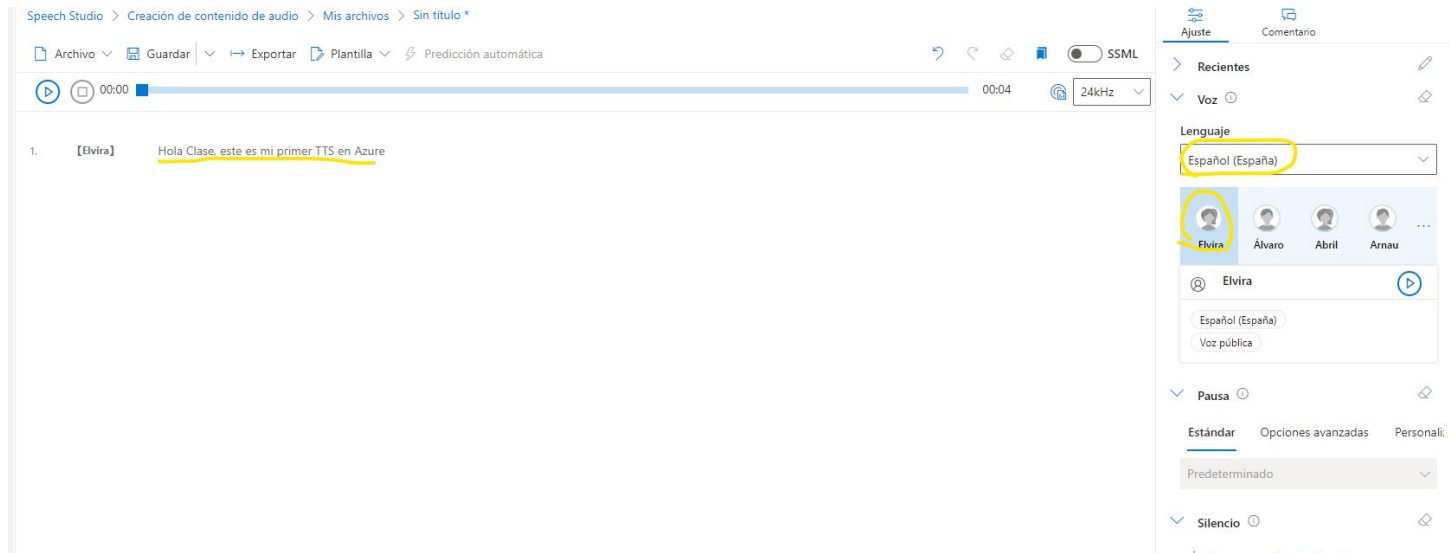
Elabore un discurso matizado ajustando el estilo, el ritmo y la pronunciación de su contenido oral.

[Iniciar un proyecto de creación de contenido de audio](#)

2. Speech Synthesis Markup Language (SSML).



2. Speech Synthesis Markup Language (SSML).



2. Speech Synthesis Markup Language (SSML).

Ahora seleccionamos la opción SSML



```
<!--ID=B7267351-473F-409D-9765-754A8EBCDE05;Version=1[{"VoiceNameToldMapItems":[{"Id":"9314d042-c519-4087-ba99-1a49d78e2d9f","Name":"Microsoft Server Speech Text to Speech Voice (es-ES, ElviraNeural)","ShortName":"es-ES-ElviraNeural","Locale":"es-ES","VoiceType":"StandardVoice"}]}-->
<!--ID=FCB40C2B-1F9F-4C26-B1A1-CF8E67BE07D1;Version=1[{"Files":{}}-->
<!--ID=5B95B1CC-2C7B-494F-B746-CF22A0E779B7;Version=1[{"Locales":{"es-ES":{"AutoApplyCustomLexiconFiles":{}}}}-->
<speak xmlns="http://www.w3.org/2001/10/synthesis" xmlns:mstts="http://www.w3.org/2001/mstts" xmlns:emo="http://www.w3.org/2009/10/emotionml" version="1.0" xml:lang="es-ES">
<voice name="es-ES-ElviraNeural">Hola Clase, este es mi primer TTS en Azure</voice>
</speak>
```

2. Speech Synthesis Markup Language (SSML).

Ahora seleccionamos la opción SSML



```
<!--ID=B7267351-473F-409D-9765-754A8EBCDE05;Version=1[{"VoiceNameToldMapItems":[{"Id":"9314d042-c519-4087-ba99-1a49d78e2d9f","Name":"Microsoft Server Speech Text to Speech Voice (es-ES, ElviraNeural)","ShortName":"es-ES-ElviraNeural","Locale":"es-ES","VoiceType":"StandardVoice"}]}-->
<!--ID=FCB40C2B-1F9F-4C26-B1A1-CF8E67BE07D1;Version=1[{"Files":{}}-->
<!--ID=5B95B1CC-2C7B-494F-B746-CF22A0E779B7;Version=1[{"Locales":{"es-ES":{"AutoApplyCustomLexiconFiles":{}}}}-->
<speak xmlns="http://www.w3.org/2001/10/synthesis" xmlns:mstts="http://www.w3.org/2001/mstts" xmlns:emo="http://www.w3.org/2009/10/emotionml" version="1.0" xml:lang="es-ES">
<voice name="es-ES-ElviraNeural">Hola Clase, este es mi primer TTS en Azure</voice>
</speak>
```

2. Speech Synthesis Markup Language (SSML).



Speak

El elemento speak contiene información, como la versión, el idioma y la definición del vocabulario de marcado. El elemento speak es el elemento raíz que se necesita para todos los documentos SSML. Debe especificar el idioma predeterminado dentro del speak elemento, independientemente de que el idioma se ajuste o no en otro lugar como, por ejemplo, dentro del elemento lang.

2. Speech Synthesis Markup Language (SSML).

Speak

Atributo	Descripción	Obligatorio u opcional
version	Indica la versión de la especificación SSML utilizada para interpretar el marcado del documento. La versión actual es "1.0".	Obligatorio
xml:lang	El idioma del documento raíz. El valor puede contener un código de idioma como <code>en</code> (inglés) o una configuración regional como <code>en-US</code> (inglés de Estados Unidos).	Obligatorio
xmlns	El URI del documento que define el vocabulario de marcado (los tipos de elementos y nombres de atributos) del documento SSML. El identificador URI actual es "http://www.w3.org/2001/10/synthesis".	Obligatorio

2. Speech Synthesis Markup Language (SSML).



VOICE NAME

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
xml:lang="es-ES">
```

```
<voice name="es-ES-ElviraNeural">
```

Hola Clase, este es mi primer TTS en Azure

```
</voice></speak>
```

2. Speech Synthesis Markup Language (SSML).

BREAK

Atributo	Descripción	Obligatorio u opcional
<code>strength</code>	La duración relativa de una pausa mediante uno de los valores siguientes: <ul style="list-style-type: none">• x-weak• weak• medium (default)• strong• x-strong	Opcional
<code>time</code>	La duración absoluta de una pausa en segundos (como <code>2s</code>) o milisegundos (como <code>500ms</code>). Los valores válidos oscilan entre 0 y 5000 milisegundos. Si establece un valor mayor que el máximo admitido, el servicio usa <code>5000ms</code> . Si establece el atributo <code>time</code> , se omite el atributo <code>strength</code> .	Opcional

2. Speech Synthesis Markup Language (SSML).

BREAK

`< speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="es-ES"> < voice name="es-ES-ElviraNeural">`

Hola Clase< break /> este es mi primer TTS en Azure

Hola Clase< break strength="medium" /> este es mi primer TTS en Azure

Hola Clase< break time="750ms" />este es mi primer TTS en Azure

`</voice> </speak>`

2. Speech Synthesis Markup Language (SSML).



mstts:silence

Usamos el elemento **mstts:silence** para insertar pausas antes o después del texto, o bien entre dos frases adyacentes.

Una de las diferencias entre **mstts:silence** y **break** es que un elemento **break** se puede insertar en cualquier parte del texto. El silencio solo funciona al principio o al final del texto de entrada o en el límite de dos frases adyacentes.

2. Speech Synthesis Markup Language (SSML).

Atributo	Descripción	Obligatorio u opcional
type	<p>Especifica dónde y cómo agregar silencio. Se admiten los siguientes tipos de silencios:</p> <p>Leading – Silencio adicional al principio del texto. El valor que establezca se agrega al silencio natural antes del inicio del texto.</p> <p>Leading-exact – Silencio al principio del texto. El valor es una longitud de silencio absoluta.</p> <p>Tailing – Silencio adicional al final del texto. El valor que establezca se agrega al silencio natural después de la última palabra.</p> <p>Tailing-exact – Silencio al final del texto. El valor es una longitud de silencio absoluta.</p> <p>Sentenceboundary – Silencio adicional entre oraciones adyacentes. La longitud real del silencio de este tipo incluye el silencio natural después de la última palabra de la oración anterior, el valor que estableció para este tipo y el silencio natural antes de la palabra inicial en la siguiente oración.</p> <p>Sentenceboundary-exact – Silencio entre oraciones adyacentes. El valor es una longitud de silencio absoluta.</p> <p>Comma-exact: silencio en la coma en formato de ancho medio o ancho completo. El valor es una longitud de silencio absoluta.</p> <p>Semicolon-exact: silencio en el punto y coma en formato de ancho medio o ancho completo. El valor es una longitud de silencio absoluta.</p> <p>Enumerationcomma-exact: silencio en la coma de enumeración en formato de ancho completo. El valor es una longitud de silencio absoluta.</p>	Obligatorio

2. Speech Synthesis Markup Language (SSML).



Un tipo de silencio absoluto (con el sufijo `-exact`) reemplaza cualquier silencio inicial o final natural. Los tipos de silencio absoluto tienen prioridad sobre el tipo no absoluto correspondiente. Por ejemplo, si establece ambos tipos `Leading` y `Leading-exact`, se aplicará el tipo `Leading-exact`. El evento `WordBoundary` tiene prioridad sobre la configuración de silencio relacionada con los signos de puntuación, como `Comma-exact`, `Semicolon-exact` o `Enumerationcomma-exact`. Cuando use a la vez el evento `WordBoundary` y la configuración de silencio mediante signos de puntuación, esta última no tendrá efecto.

Value

La duración de una pausa en segundos (como `2s`) o milisegundos (como `500ms`). Los valores válidos oscilan entre 0 y 5000 milisegundos. Si establece un valor mayor que el máximo admitido, el servicio usa `5000ms`.

Obligatorio

2. Speech Synthesis Markup Language (SSML).



mstts:silence Ejemplo: Aplicar 200 ms de silencio entre dos frases.

```
< speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="es-ES"> < voice  
name="es-ES-ElviraNeural">
```

```
<mstts:silence type="Sentenceboundary" value="200ms"/>
```

La perseverancia es la clave del éxito; incluso en los momentos más difíciles, sigue adelante con determinación.

La verdadera grandeza no se mide por la riqueza material, sino por la bondad y compasión que mostramos hacia los demás


```
</voice> </speak>
```

2. Speech Synthesis Markup Language (SSML).



Los elementos **p** y **s** se utilizan para indicar párrafos y frases, respectivamente. En ausencia de estos elementos, el servicio de Voz determina automáticamente la estructura del documento SSML.

2. Speech Synthesis Markup Language (SSML).



```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="es-ES"> <voice name="es-ES-ElviraNeural">
```

<p> Fragmento de un escrito con unidad temática, que queda diferenciado del resto de fragmentos por un punto y aparte y generalmente también por llevar letra mayúscula inicial y un espacio en blanco en el margen izquierdo de alineación del texto principal de la primera línea.<s>Unidad lingüística que consta de sujeto y predicado, implícitos o explícitos; sintácticamente, puede ser simple (independiente) o compuesta (coordinada o subordinada).</s></p>

```
</voice> </speak>
```

2. Speech Synthesis Markup Language (SSML).

Podemos utilizar múltiples voces y aplicar efectos.

effect Procesador de efectos de audio que se usa para optimizar la calidad de la salida de voz sintetizada para escenarios específicos en dispositivos.

En algunos escenarios de entornos de producción, la experiencia auditiva puede degradarse debido a la distorsión de la reproducción en ciertos dispositivos. Por ejemplo, la voz sintetizada de un altavoz de coche puede sonar apagada debido a factores ambientales como la respuesta del hablante, la reverberación de la sala y el ruido de fondo. Es posible que el pasajero tenga que subir el volumen para oír con más claridad. Para evitar las operaciones manuales en un escenario así, el procesador de efectos de audio puede hacer que el sonido sea más claro compensando la distorsión de la reproducción.

Se admiten los valores siguientes:

eq_car – Optimice la experiencia auditiva al proporcionar voz de alta calidad en coches, autobuses y otros vehículos cerrados.

eq_telecomhp8k – Optimice la experiencia auditiva para la voz de banda estrecha en escenarios de telecomunicaciones o telefonía. Debe usar una frecuencia de muestreo de 8 kHz. Si la frecuencia de muestreo no es de 8 kHz, la calidad auditiva de la voz de salida no se optimizará.

Si falta el valor o no es válido, este atributo se omitirá y no se aplicará ningún efecto.

2. Speech Synthesis Markup Language (SSML).



Ejemplo con múltiples voces.

```
<speech version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">  
  <voice name="en-US-JennyNeural">  
    Good morning!  
  </voice>  
  <voice name="en-US-ChristopherNeural">  
    Good morning to you too Jenny!  
  </voice>  
</speech>
```


2. Speech Synthesis Markup Language (SSML).



Ejemplo con efecto de audio.

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">  
  <voice name="en-US-JennyNeural" effect="eq_car">  
    This is the text that is spoken.  
  </voice>  
</speak>
```

2. Speech Synthesis Markup Language (SSML).



mstts:express-as

Estilo de habla específico de voz. Puede expresar emociones como alegría, empatía y calma. También puede optimizar la voz para distintos escenarios, como el servicio de atención al cliente, las noticias y el asistente de voz.

2. Speech Synthesis Markup Language (SSML).

mstts:express-as

Estilo	Descripción
<code>style="advertisement_upbeat"</code>	Expresa un tono emocionado y con mucha energía para promover un producto o servicio.
<code>style="affectionate"</code>	Expresa un tono cálido y afectuoso, agudo y con una energía vocal alta. El hablante está en un estado que atrae la atención de su interlocutor. La personalidad del hablante suele ser simpática por naturaleza.
<code>style="angry"</code>	Expresa un tono enfadado y molesto.
<code>style="assistant"</code>	Expresa un tono cálido y relajado para los asistentes digitales.
<code>style="calm"</code>	Expresa una actitud interesante, recolectada y compuesta al hablar. El tono, la intensidad y la prosodia son más uniformes en comparación con otros tipos de voz.
<code>style="chat"</code>	Expresa un tono casual y relajado.
<code>style="cheerful"</code>	Expresa un tono positivo y feliz.
<code>style="customerservice"</code>	Expresa un tono amistoso y servicial para la asistencia al cliente.

2. Speech Synthesis Markup Language (SSML).

mstts:express-as

style="customerservice"	Expresa un tono amistoso y servicial para la asistencia al cliente.
style="depressed"	Expresa un tono melancólico y pesimista, con un tono más grave y energía vocal.
style="disgruntled"	Expresa desdén y un tono de queja. La voz de esta emoción muestra desagrado y desprecio.
style="documentary-narration"	Narra documentales en un estilo relajado, interesante e informativo adecuado para doblaje documentales, comentarios expertos y contenido similar.
style="embarrassed"	Expresa un tono vacilante e inseguro cuando quien habla se siente incómodo.
style="empathetic"	Expresa un sentimiento de cuidado y comprensión.
style="envious"	Expresa un tono de admiración cuando desea algo que otra persona tiene.
style="excited"	Expresa un tono optimista y esperanzador. Suena como si algo grande estuviera sucediendo y el orador estuviera feliz por ello.
style="fearful"	Expresa un tono asustado y nervioso, agudo, con energía vocal alta y más velocidad. El hablante está en un estado de tensión y desasosiego.
style="friendly"	Expresa un tono agradable, acogedor y cálido. Suena sincero y cariñoso.

2. Speech Synthesis Markup Language (SSML).

mstts:express-as

style="gentle"	Expresa un tono dulce, educado y agradable, con un tono más grave y energía vocal.
style="hopeful"	Expresa un tono cálido y de deseo. Parece que algo bueno le pasará al orador.
style="lyrical"	Expresa emociones de manera melódica y sentimental.
style="narration-professional"	Expresa un tono objetivo y profesional para la lectura de contenido.
style="narration-relaxed"	Expresa un tono relajante y melodioso para la lectura de contenidos.
style="newscast"	Expresa un tono formal y profesional para narrar noticias.
style="newscast-casual"	Expresa un tono versátil e informal para la difusión de noticias generales.
style="newscast-formal"	Expresa un tono formal, seguro y autoritario para la difusión de noticias.
style="poetry-reading"	Expresa un tono emocional y rítmico mientras lee un poema.
style="sad"	Expresa un tono afligido.

2. Speech Synthesis Markup Language (SSML).

mstts:express-as

<code>style="serious"</code>	Expresa un tono estricto e imponente. A menudo, el hablante suena rígido y mucho menos relajado, con una cadencia firme.
<code>style="shouting"</code>	Expresa un tono que suena como si la voz estuviera distante o en otro lugar y haciendo un esfuerzo para que se le oiga claramente.
<code>style="sports_commentary"</code>	Expresa un tono relajado e interesado para retransmitir un acontecimiento deportivo.
<code>style="sports_commentary_excited"</code>	Expresa un tono intenso y enérgico para transmitir momentos emocionantes en un evento deportivo.
<code>style="whispering"</code>	Expresa un tono suave que intenta emitir un sonido tranquilo y apacible.
<code>style="terrified"</code>	Expresa un tono asustado, con un ritmo más rápido y una voz más temblorosa. Suena como si el orador estuviera en un estado inestable y frenético.
<code>style="unfriendly"</code>	Expresa un tono frío e indiferente.

2. Speech Synthesis Markup Language (SSML).

mstts:express-as En el siguiente ejemplo de SSML se usa el elemento con un grado de estilo sad de 2.

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
```

```
<voice name="en-US-JennyNeural" effect="eq_car">
```

```
<mstts:express-as style="sad" styledegree="2">
```

This is the text that is spoken.

```
</mstts:express-as>
```

```
</voice> </speak>
```

IES ADA2020S

2. Speech Synthesis Markup Language (SSML).



mstts:express-as No todas las voces aceptan los estilos, para más información:

<https://learn.microsoft.com/es-es/azure/ai-services/speech-service/language-support?tabs=tts>

2. Speech Synthesis Markup Language (SSML).



Podemos construir SSML multidioma:

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
xmlns:mstts="https://www.w3.org/2001/mstts" xml:lang="en-US">
  <voice name="en-US-JennyMultilingualNeural">
    <lang xml:lang="es-MX">
      ¡Esperamos trabajar con usted!
    </lang>
    <lang xml:lang="en-US">
      We look forward to working with you!
    </lang>
    <lang xml:lang="fr-FR">
      Nous avons hâte de travailler avec vous!
    </lang>
  </voice>
</speak>
```

IES Abastos

2. Speech Synthesis Markup Language (SSML).



prosody: Para especificar los cambios en el tono, la curva melódica, el rango, la velocidad y el volumen de la salida de texto a voz.

2. Speech Synthesis Markup Language (SSML).

prosody

contour

La curva melódica representa los cambios en el tono. Estos cambios se representan como una matriz de objetivos en posiciones de tiempo específicas en la salida de voz. Los conjuntos de pares de parámetros definen cada destino. Por ejemplo:

Opcional

```
<prosody contour="(0%,+20Hz) (10%,-2st) (40%,+10Hz)">
```

El primer valor de cada conjunto de parámetros especifica la ubicación del cambio de tono como porcentaje de la duración del texto. El segundo valor especifica la cantidad que se va subir o bajar el tono, mediante un valor relativo o un valor de enumeración para el tono (consulte `pitch`).

2. Speech Synthesis Markup Language (SSML).

prosody

pitch

Indica el tono de la línea de referencia del texto. Los cambios de tono se pueden aplicar en el nivel de oración. Los cambios de tono deben estar comprendidos entre la 0,5 y 1,5 veces el audio original. Puede expresar el tono como:

Un valor absoluto: expresado como un número seguido de "Hz" (Hercios). Por ejemplo, `<prosody pitch="600Hz">some text</prosody>`.

Valor relativo:

Como número relativo: expresado como un número precedido por "+" o "-" y seguido por "Hz" o "st", que especifica una cantidad para cambiar el tono. Por ejemplo: `<prosody pitch="+80Hz">some text</prosody>` o `<prosody pitch="-2st">some text</prosody>`. La "st" indica que la unidad de cambio es un semitono, que es la mitad de un tono (medio paso) en la escala diatónica estándar.

Como porcentaje: expresado como un número precedido por "+" (opcionalmente) o "-" y seguido de "%", que indica el cambio relativo. Por ejemplo: `<prosody pitch="50%">some text</prosody>` o `<prosody pitch="-50%">some text</prosody>`.

Un valor constante:

x-low

low

medio

high

x-high

default

2. Speech Synthesis Markup Language (SSML).

prosody

range Un valor que representa el rango del tono para el texto. Puede expresar el elemento **range** mediante los mismos valores absolutos, valores relativos o valores de enumeración usados para describir el elemento **pitch**.

rate Indica la velocidad de habla del texto. La velocidad de habla se puede aplicar tanto en el nivel de palabra como en el de frase. Los cambios de velocidad deben estar comprendidos entre **0.5** y **2** veces el audio original. Puede expresar **rate** como:

Valor relativo:

Como número relativo: expresado como un número que actúa como multiplicador del valor predeterminado. Por ejemplo, un valor de **1** no da como resultado ningún cambio en la velocidad original. Un valor de **0.5** da como resultado la mitad de la velocidad original. Un valor de **2** da como resultado el doble de la velocidad original.

Como porcentaje: expresado como un número precedido por "+" (opcionalmente) o "-" y seguido de "%", que indica el cambio relativo. Por ejemplo: `<prosody rate="50%">some text</prosody>` o `<prosody rate="-50%">some text</prosody>`.

Un valor constante:

x-slow

lento

medio

fast

x-fast

default

2. Speech Synthesis Markup Language (SSML).

prosody

volume

Indica el nivel de volumen de la voz. Los cambios de volumen se pueden aplicar en el nivel de oración. Puede expresar el volumen como:

Un valor absoluto, expresado como un número en el rango de `0.0` a `100.0`, de *más silencioso* a *más alto*, como `75`. El valor predeterminado es `100.0`.

Valor relativo:

Como número relativo: expresado como un número precedido por "+" o "-" que especifica una cantidad para cambiar el volumen. Algunos ejemplos son `+10` o `-5.5`.

Como porcentaje: expresado como un número precedido por "+" (opcionalmente) o "-" y seguido de "%", que indica el cambio relativo. Por ejemplo: `<prosody volume="50%">some text</prosody>` o `<prosody volume="+3%">some text</prosody>`.

Un valor constante:

`silent`

`x-soft`

`soft`

`medio`

`loud`

`x-loud`

`default`

2. Speech Synthesis Markup Language (SSML).



Ejemplo de cambio de la velocidad de habla:

Este fragmento de código SSML muestra cómo se usa el atributo `rate` para aumentar un 30 % la velocidad de habla a partir de la velocidad predeterminada.

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  <voice name="en-US-JennyNeural">
    <prosody rate="+30.00%">
      Enjoy using text to speech.
    </prosody>
  </voice>
</speak>
```

2. Speech Synthesis Markup Language (SSML).



Ejemplo de cambio de volumen

Este fragmento de código SSML muestra cómo se usa el atributo `volume` para aumentar un 20 % el volumen a partir del volumen predeterminado.

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  <voice name="en-US-JennyNeural">
    <prosody volume="+20.00%">
      Enjoy using text to speech.
    </prosody>
  </voice>
</speak>
```


2. Speech Synthesis Markup Language (SSML).



Ejemplo de cambio de tono

Este fragmento de código SSML muestra cómo se usa el atributo `pitch` para que la voz hable en un tono alto.

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  <voice name="en-US-JennyNeural">
    Welcome to <prosody pitch="high">Enjoy using text to speech.</prosody>
  </voice>
</speak>
```

2. Speech Synthesis Markup Language (SSML).



Ejemplo de cambio de curva melódica

Este fragmento de código SSML muestra cómo se usa el atributo `contour` para cambiar el contorno.

```
<speech version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  <voice name="en-US-JennyNeural">
    <prosody contour="(60%,-60%) (100%,+80%)" >
      Were you the only person in the room?
    </prosody>
  </voice>
</speech>
```

2. Speech Synthesis Markup Language (SSML).

emphasis Para agregar o quitar el énfasis a nivel de palabra del texto. El ajuste del énfasis de nivel de palabra solo está disponible para estas voces neuronales: en-US-GuyNeural, en-US-DavisNeural y en-US-JaneNeural.

Atributo	Descripción	Obligatorio u opcional
<code>level</code>	<p>Indica la intensidad del énfasis que se va a aplicar:</p> <ul style="list-style-type: none">• <code>reduced</code>• <code>none</code>• <code>moderate</code>• <code>strong</code> <p>Quando no se especifica el atributo <code>level</code>, el nivel predeterminado es <code>moderate</code>. Para más información sobre cada atributo, consulte el elemento de énfasis.</p>	Opcionales

2. Speech Synthesis Markup Language (SSML).

emphasis

```
<speech version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
xmlns:mstts="https://www.w3.org/2001/mstts" xml:lang="en-US">
  <voice name="en-US-GuyNeural">
    I can help you join your <emphasis level="moderate">meetings</emphasis> fast.
  </voice>
</speech>
```

2. Speech Synthesis Markup Language (SSML).

mstts:audioduration

Use este elemento para ayudar a sincronizar el tiempo de finalización de la salida de audio. La duración del audio se puede reducir o aumentar entre 0.5 y 2 veces la velocidad del audio original.

Atributo	Descripción
value	<p>La duración solicitada del audio de salida en segundos, (como 2s) o en milisegundos (como 2000ms).</p> <p>Este valor debe estar comprendido entre 0.5 y 2 veces el audio original sin ninguna otra configuración de velocidad. Por ejemplo, si la duración solicitada del audio es 30s, el audio original debe estar entre 15 y 60 segundos. Si establece un valor fuera de estos límites, la duración se establece según el múltiplo mínimo o máximo respectivo.</p> <p>Dada la duración del audio de salida solicitada, el servicio de Voz ajusta la velocidad de habla según corresponda. Use la API de lista de voces y compruebe el atributo WordsPerMinute para averiguar la velocidad de habla de la voz neuronal que usa. Puede dividir el número de palabras del texto de entrada por el valor del atributo WordsPerMinute para obtener la duración aproximada del audio de salida original. El audio de salida suena más natural cuando establece la duración del audio más cercana a la duración estimada.</p>

2. Speech Synthesis Markup Language (SSML).



mstts:audioduration

```
< speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xmlns:mstts="http://www.w3.org/2001/mstts"
xml:lang="en-US">
< voice name="en-US-JennyNeural">
<mstts:audioduration value="20s"/>
```

If we're home schooling, the best we can do is roll with what each day brings and try to have fun along the way.
A good place to start is by trying out the slew of educational apps that are helping children stay happy and smash their schooling at the same time.

```
</voice>
</speak>
```

2. Speech Synthesis Markup Language (SSML).

Say-as Indica el tipo de contenido (por ejemplo, un número o una fecha) del texto del elemento. Este elemento proporciona instrucciones al motor de síntesis de voz sobre cómo pronunciar el texto.

Atributo	Descripción	Obligatorio u opcional
<code>interpret-as</code>	Indica el tipo de contenido del texto de un elemento. Para obtener una lista de tipos, consulte la tabla siguiente.	Obligatorio
<code>format</code>	Proporciona información adicional sobre el formato preciso del texto del elemento para los tipos de contenido que pueden tener formatos ambiguos. SSML define formatos para los tipos de contenido que los usan. Consulte la tabla siguiente.	Opcional
<code>detail</code>	Indica el nivel de detalle con que se va a hablar. Por ejemplo, este atributo puede solicitar que el motor de síntesis de voz pronuncie signos de puntuación. No hay valores estándar definidos para <code>detail</code> .	Opcional

2. Speech Synthesis Markup Language (SSML).

Say-as

characters,
spell-out

El texto se pronuncia como letras individuales (deletreadas). El motor de síntesis de voz pronuncia:

```
<say-as interpret-as="characters">test</say-as>
```

Como "T E S T".

cardinal,
number

Ninguno

El texto se pronuncia como un número cardinal. El motor de síntesis de voz pronuncia:

```
There are <say-as interpret-as="cardinal">10</say-as> options
```

Como "Hay diez opciones."

ordinal

None

El texto se pronuncia como un número ordinal. El motor de síntesis de voz pronuncia:

```
Select the <say-as interpret-as="ordinal">3rd</say-as> option
```

Como "Seleccione la tercera opción".

2. Speech Synthesis Markup Language (SSML).

Say-as

<code>number_digit</code>	Ninguno	<p>El texto se pronuncia como una secuencia de dígitos individuales. El motor de síntesis de voz pronuncia:</p> <pre><say-as interpret-as="number_digit">123456789</say-as></pre> <p>Como "1 2 3 4 5 6 7 8 9".</p>
<code>fraction</code>	None	<p>El texto se pronuncia como un número fraccionario. El motor de síntesis de voz pronuncia:</p> <pre><say-as interpret-as="fraction">3/8</say-as> of an inch</pre> <p>Como "tres octavos de pulgada".</p>
<code>date</code>	dmy, mdy, ymd, ydm, ym, my, md, dm, d, m, y	<p>El texto se pronuncia como una fecha. El atributo <code>format</code> especifica el formato de la fecha (<i>d=día, m=mes, y=año</i>). El motor de síntesis de voz pronuncia:</p> <pre>Today is <say-as interpret-as="date">10-12-2016</say-as></pre> <p>Como "Hoy es el duodécimo de octubre dos mil dieciséis."</p>

2. Speech Synthesis Markup Language (SSML).

Say-as

time

hms12, hms24

El texto se pronuncia como una hora. El atributo `format` especifica si la hora se especifica mediante un reloj de 12 horas (hms12) o de 24 horas (hms24). Use un signo de dos puntos para separar los números que representan las horas, los minutos y los segundos. Estos son algunos ejemplos de horas válidas: 12:35, 1:14:32, 08:15 y 02:50:45. El motor de síntesis de voz pronuncia:

```
The train departs at <say-as interpret-as="time" format="hms12">4:00am</say-as>
```

Como "El tren sale a las cuatro A. M.".

duration

hms, hm, ms

El texto se pronuncia como una duración. El atributo `format` especifica el formato de la duración (*h=hour, m=minute y s=second*). El motor de síntesis de voz pronuncia:

```
<say-as interpret-as="duration">01:18:30</say-as>
```

Como "una hora, dieciocho minutos y treinta segundos".

Pronuncia:

```
<say-as interpret-as="duration" format="ms">01:18</say-as>
```

2. Speech Synthesis Markup Language (SSML).

Say-as

telephone

None

El texto se pronuncia como un número de teléfono. El motor de síntesis de voz pronuncia:

```
The number is <say-as interpret-as="telephone">(888) 555-1212</say-as>
```

Como "Mi número es el código de área ocho ocho ocho cinco cinco cinco uno dos uno dos".

currency

None

El texto se pronuncia como una moneda. El motor de síntesis de voz pronuncia:

```
<say-as interpret-as="currency">99.9 USD</say-as>
```

Como "noventa y nueve dólares y noventa centavos estadounidenses".

address

None

El texto se pronuncia como una dirección. El motor de síntesis de voz pronuncia:

```
I'm at <say-as interpret-as="address">150th CT NE, Redmond, WA</say-as>
```

Como "Estoy en el número 150 de Court North East Redmond, en Washington".

2. Speech Synthesis Markup Language (SSML).

Say-as

name

None

El texto se pronuncia como el nombre de una persona. El motor de síntesis de voz pronuncia:

```
<say-as interpret-as="name">ED</say-as>
```

Como [æd].

En los nombres chinos, algunos caracteres se pronuncian de forma diferente cuando aparecen en un nombre de familia. Por ejemplo, el motor de síntesis de voz dice 仇 en

```
<say-as interpret-as="name">仇先生</say-as>
```

Como [qiú] en lugar de [chóu].

2. Speech Synthesis Markup Language (SSML).

Say-as

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
```

```
  <voice name="en-US-JennyNeural">
```

```
    <p>
```

```
      Your <say-as interpret-as="ordinal"> 1st </say-as> request was for <say-as interpret-as="cardinal"> 1 </say-as> room
```

```
      on <say-as interpret-as="date" format="mdy"> 10/19/2010 </say-as>, with early arrival at <say-as interpret-as="time"
      format="hms12"> 12:35pm </say-as>.
```

```
    </p>
```

```
  </voice>
```

```
</speak>
```

ES Abastos