



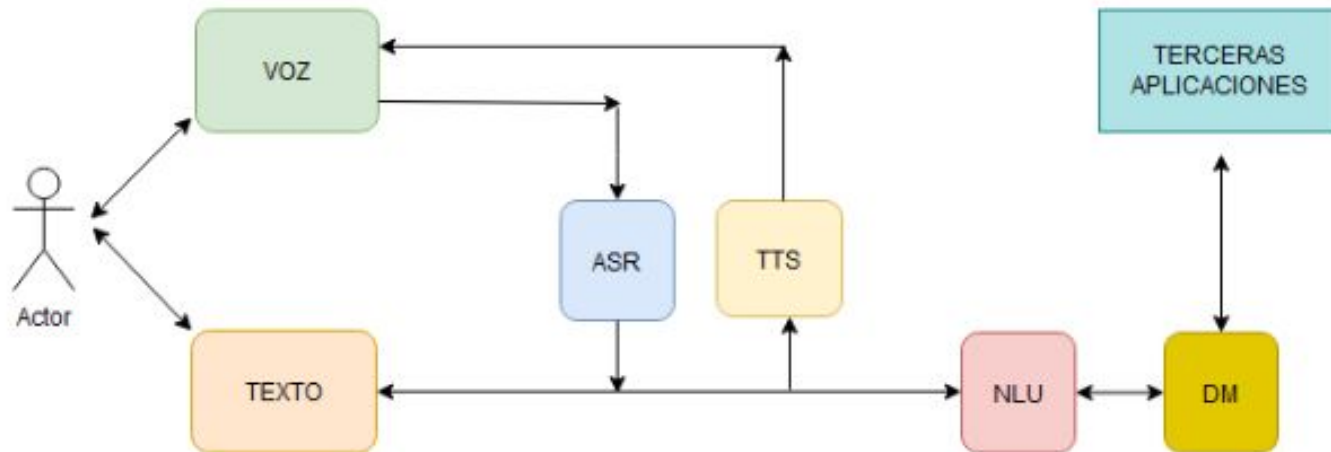
UT.9 PROCESAMIENTO DEL LENGUAJE NATURAL(II Parte): VOZ A TEXTO

Bloques de la unidad:



1. Text to Speech (TTS).
2. Speech Synthesis Markup Language (SSML).
3. Laboratorio 1: SSML.
4. Laboratorio 2: Control de costes.
5. Voz a texto (ASR).
6. Traductor.
7. Laboratorio 3 Transcribir, traducir y reproducir audios.

5. Voz a texto (ASR)



5. Voz a texto (ASR)



Nos conectamos a Azure y seleccionamos Speech Studio.

Hay que tener en cuenta que hay dos tipos de utilización de ASR:

- En tiempo real.
- Por lote.

5. Voz a texto (ASR) - Tiempo real.



En tiempo real: Enviamos la voz en tiempo real a un SDK, API o End point.

El mayor problema es la sincronización de las transcripciones.

Azure nos va a devolver un desplazamiento y la duración de la voz reconocida.

5. Voz a texto (ASR) - Tiempo real.



Desplazamiento: es el desplazamiento en la secuencia de audio que se reconoce, expresado en forma de duración. El desplazamiento se mide en tics, empezando por el tic 0 (cero), que está asociado al primer byte de audio que procesa el SDK. Es decir, el desplazamiento comienza cuando se inicia el reconocimiento, ya que es cuando el SDK comienza a procesar la secuencia de audio. Un tic representa cien nanosegundos o una diezmillonésima parte de un segundo.

Duración: es la duración de la expresión que se reconoce. La duración en tics no incluye el silencio final ni el inicial.

5. Voz a texto (ASR) - Tiempo real.



Vamos a disponer de 2 eventos:

- RECOGNIZING: Estamos procesando la transcripción.
- RECOGNIZED: Hemos finalizando la transcripción.

En la siguiente diapositiva vamos a ver un ejemplo con la siguiente frase: “Le damos la bienvenida al curso de matemáticas aplicadas 201”

5. Voz a texto (ASR) - Tiempo real.

Evento	Texto	Desplazamiento (en tics)	Duración (en tics)
RECOGNIZING	bienvenido	17 000 000	5000000
RECOGNIZING	Le damos la	17 000 000	6 400 000
RECOGNIZING	Le damos la bienvenida al	17 000 000	13 600 000
RECOGNIZING	Le damos la bienvenida al curso de	17 000 000	17 200 000
RECOGNIZING	Le damos la bienvenida al curso de matemáticas	17 000 000	23 700 000
RECOGNIZING	Le damos la bienvenida al curso de matemáticas aplicadas 2	17 000 000	26 700 000
RECOGNIZING	Le damos la bienvenida al curso de matemáticas aplicadas 201	17 000 000	33 400 000
RECOGNIZED	Le damos la bienvenida al curso de matemáticas aplicadas 201.	17 000 000	34 500 000

La duración total de la primera expresión fue de 3,45 segundos.

Duración total: 3,45

al y Big Data/ Modelos de Inteligencia Artificial

5. Voz a texto (ASR) - Tiempo real.



Evento	Texto	Desplazamiento (en tics)	Duración (en tics)
RECOGNIZING	Aceptar	71 500 000	3 100 000
RECOGNIZING	Muy bien	71 500 000	10 300 000
RECOGNIZING	Muy bien, ahora	71 500 000	14 700 000
RECOGNIZING	Muy bien, ahora comencemos.	71 500 000	18 500 000
RECOGNIZED	Muy bien, ahora comencemos.	71 500 000	20 600 000

¿ Cual es la duración ?

5. Voz a texto (ASR) - Tiempo real.

Dentro de Speech Studio seleccionamos “Voz en tiempo real” y grabamos un audio con el micrófono. Los que no tengáis micro podeis subir un archivo de audio. Es muy importante seleccionar la pestaña JSON

Usar con el recurso

☒ Reconozco que esta aplicación usa el recurso Estenombretienequeserunico y que implicará el uso de mi cuenta. [Elige otro recurso](#)

Elegir un idioma

Español (España) ▼

[Mostrar opciones avanzadas](#)

Elegir los archivos de audio

Arrastrar y colocar archivos de audio aquí o
[Examinar archivos...](#)

O grabar audio con un micrófono


Resultados de pruebas

Nombre de archivo: --	Lenguaje: --	Formato de salida: --
Punto de conexión personalizado: --	Lista de frases: --	
<p>▶ 00:00 ○ 00:00s ↓</p>		


5. Voz a texto (ASR) - Tiempo real.

Ahora vamos a analizar el JSON, queremos analizar la duración, el desplazamiento y la gestión de eventos.

Elegir los archivos de audio






Arrastrar y colocar archivos de audio aquí o
[Examinar archivos...](#)



O grabar audio con un micrófono

Archivos de audio

 20240200_115511.wav

Resultados de pruebas


Nombre de archivo: 20240200_115511.wav

Lenguaje: Español (España)


Formato de salida: Detallado

Punto de conexión personalizado: [Ninguno]

Lista de frases: Desactivado

 00:00

00:10s



Texto

JSON

```
1  {
2
3    "Id": "8d99305883e64b67a47861215cd9db55",
4    "RecognitionStatus": 0,
5    "Offset": 11900000,
6    "Duration": 7760000,
7    "Channel": 0,
8    "DisplayText": "Hola, estamos en la clase de modelos y esto es una prueba de una",
9    "NBest": [
10     {
11       "Confidence": 0.9337882,
12       "Lexical": "hola estamos en la clase de modelos y esto es una prueba de",
13       "ITN": "hola estamos en la clase de modelos y esto es una prueba de una",
14       "MaskedITN": "hola estamos en la clase de modelos y esto es una prueba de",
15       "Display": "Hola, estamos en la clase de modelos y esto es una prueba de",
16       "Words": [
17         {
18           "Text": "hola",
19           "Offset": 11900000,
20           "Duration": 1000000,
21           "Confidence": 0.9337882,
22           "Lexical": "hola",
23           "ITN": "hola",
24           "MaskedITN": "hola",
25           "Display": "hola",
26           "Words": ["hola"]
27         },
28         {
29           "Text": "estamos",
30           "Offset": 12900000,
31           "Duration": 1000000,
32           "Confidence": 0.9337882,
33           "Lexical": "estamos",
34           "ITN": "estamos",
35           "MaskedITN": "estamos",
36           "Display": "estamos",
37           "Words": ["estamos"]
38         },
39         {
40           "Text": "en",
41           "Offset": 13900000,
42           "Duration": 1000000,
43           "Confidence": 0.9337882,
44           "Lexical": "en",
45           "ITN": "en",
46           "MaskedITN": "en",
47           "Display": "en",
48           "Words": ["en"]
49         },
50         {
51           "Text": "la",
52           "Offset": 14900000,
53           "Duration": 1000000,
54           "Confidence": 0.9337882,
55           "Lexical": "la",
56           "ITN": "la",
57           "MaskedITN": "la",
58           "Display": "la",
59           "Words": ["la"]
60         },
61         {
62           "Text": "clase",
63           "Offset": 15900000,
64           "Duration": 1000000,
65           "Confidence": 0.9337882,
66           "Lexical": "clase",
67           "ITN": "clase",
68           "MaskedITN": "clase",
69           "Display": "clase",
70           "Words": ["clase"]
71         },
72         {
73           "Text": "de",
74           "Offset": 16900000,
75           "Duration": 1000000,
76           "Confidence": 0.9337882,
77           "Lexical": "de",
78           "ITN": "de",
79           "MaskedITN": "de",
80           "Display": "de",
81           "Words": ["de"]
82         },
83         {
84           "Text": "modelos",
85           "Offset": 17900000,
86           "Duration": 1000000,
87           "Confidence": 0.9337882,
88           "Lexical": "modelos",
89           "ITN": "modelos",
90           "MaskedITN": "modelos",
91           "Display": "modelos",
92           "Words": ["modelos"]
93         },
94         {
95           "Text": "y",
96           "Offset": 18900000,
97           "Duration": 1000000,
98           "Confidence": 0.9337882,
99           "Lexical": "y",
100          "ITN": "y",
101          "MaskedITN": "y",
102          "Display": "y",
103          "Words": ["y"]
104        },
105        {
106          "Text": "esto",
107          "Offset": 19900000,
108          "Duration": 1000000,
109          "Confidence": 0.9337882,
110          "Lexical": "esto",
111          "ITN": "esto",
112          "MaskedITN": "esto",
113          "Display": "esto",
114          "Words": ["esto"]
115        },
116        {
117          "Text": "es",
118          "Offset": 20900000,
119          "Duration": 1000000,
120          "Confidence": 0.9337882,
121          "Lexical": "es",
122          "ITN": "es",
123          "MaskedITN": "es",
124          "Display": "es",
125          "Words": ["es"]
126        },
127        {
128          "Text": "una",
129          "Offset": 21900000,
130          "Duration": 1000000,
131          "Confidence": 0.9337882,
132          "Lexical": "una",
133          "ITN": "una",
134          "MaskedITN": "una",
135          "Display": "una",
136          "Words": ["una"]
137        },
138        {
139          "Text": "prueba",
140          "Offset": 22900000,
141          "Duration": 1000000,
142          "Confidence": 0.9337882,
143          "Lexical": "prueba",
144          "ITN": "prueba",
145          "MaskedITN": "prueba",
146          "Display": "prueba",
147          "Words": ["prueba"]
148        },
149        {
150          "Text": "de",
151          "Offset": 23900000,
152          "Duration": 1000000,
153          "Confidence": 0.9337882,
154          "Lexical": "de",
155          "ITN": "de",
156          "MaskedITN": "de",
157          "Display": "de",
158          "Words": ["de"]
159        },
160        {
161          "Text": "una",
162          "Offset": 24900000,
163          "Duration": 1000000,
164          "Confidence": 0.9337882,
165          "Lexical": "una",
166          "ITN": "una",
167          "MaskedITN": "una",
168          "Display": "una",
169          "Words": ["una"]
170        },
171        {
172          "Text": "prueba",
173          "Offset": 25900000,
174          "Duration": 1000000,
175          "Confidence": 0.9337882,
176          "Lexical": "prueba",
177          "ITN": "prueba",
178          "MaskedITN": "prueba",
179          "Display": "prueba",
180          "Words": ["prueba"]
181        },
182        {
183          "Text": "de",
184          "Offset": 26900000,
185          "Duration": 1000000,
186          "Confidence": 0.9337882,
187          "Lexical": "de",
188          "ITN": "de",
189          "MaskedITN": "de",
190          "Display": "de",
191          "Words": ["de"]
192        },
193        {
194          "Text": "una",
195          "Offset": 27900000,
196          "Duration": 1000000,
197          "Confidence": 0.9337882,
198          "Lexical": "una",
199          "ITN": "una",
200          "MaskedITN": "una",
201          "Display": "una",
202          "Words": ["una"]
203        },
204        {
205          "Text": "prueba",
206          "Offset": 28900000,
207          "Duration": 1000000,
208          "Confidence": 0.9337882,
209          "Lexical": "prueba",
210          "ITN": "prueba",
211          "MaskedITN": "prueba",
212          "Display": "prueba",
213          "Words": ["prueba"]
214        },
215        {
216          "Text": "de",
217          "Offset": 29900000,
218          "Duration": 1000000,
219          "Confidence": 0.9337882,
220          "Lexical": "de",
221          "ITN": "de",
222          "MaskedITN": "de",
223          "Display": "de",
224          "Words": ["de"]
225        },
226        {
227          "Text": "una",
228          "Offset": 30900000,
229          "Duration": 1000000,
230          "Confidence": 0.9337882,
231          "Lexical": "una",
232          "ITN": "una",
233          "MaskedITN": "una",
234          "Display": "una",
235          "Words": ["una"]
236        },
237        {
238          "Text": "prueba",
239          "Offset": 31900000,
240          "Duration": 1000000,
241          "Confidence": 0.9337882,
242          "Lexical": "prueba",
243          "ITN": "prueba",
244          "MaskedITN": "prueba",
245          "Display": "prueba",
246          "Words": ["prueba"]
247        },
248        {
249          "Text": "de",
250          "Offset": 32900000,
251          "Duration": 1000000,
252          "Confidence": 0.9337882,
253          "Lexical": "de",
254          "ITN": "de",
255          "MaskedITN": "de",
256          "Display": "de",
257          "Words": ["de"]
258        },
259        {
260          "Text": "una",
261          "Offset": 33900000,
262          "Duration": 1000000,
263          "Confidence": 0.9337882,
264          "Lexical": "una",
265          "ITN": "una",
266          "MaskedITN": "una",
267          "Display": "una",
268          "Words": ["una"]
269        },
270        {
271          "Text": "prueba",
272          "Offset": 34900000,
273          "Duration": 1000000,
274          "Confidence": 0.9337882,
275          "Lexical": "prueba",
276          "ITN": "prueba",
277          "MaskedITN": "prueba",
278          "Display": "prueba",
279          "Words": ["prueba"]
280        },
281        {
282          "Text": "de",
283          "Offset": 35900000,
284          "Duration": 1000000,
285          "Confidence": 0.9337882,
286          "Lexical": "de",
287          "ITN": "de",
288          "MaskedITN": "de",
289          "Display": "de",
290          "Words": ["de"]
291        },
292        {
293          "Text": "una",
294          "Offset": 36900000,
295          "Duration": 1000000,
296          "Confidence": 0.9337882,
297          "Lexical": "una",
298          "ITN": "una",
299          "MaskedITN": "una",
300          "Display": "una",
301          "Words": ["una"]
302        },
303        {
304          "Text": "prueba",
305          "Offset": 37900000,
306          "Duration": 1000000,
307          "Confidence": 0.9337882,
308          "Lexical": "prueba",
309          "ITN": "prueba",
310          "MaskedITN": "prueba",
311          "Display": "prueba",
312          "Words": ["prueba"]
313        },
314        {
315          "Text": "de",
316          "Offset": 38900000,
317          "Duration": 1000000,
318          "Confidence": 0.9337882,
319          "Lexical": "de",
320          "ITN": "de",
321          "MaskedITN": "de",
322          "Display": "de",
323          "Words": ["de"]
324        },
325        {
326          "Text": "una",
327          "Offset": 39900000,
328          "Duration": 1000000,
329          "Confidence": 0.9337882,
330          "Lexical": "una",
331          "ITN": "una",
332          "MaskedITN": "una",
333          "Display": "una",
334          "Words": ["una"]
335        },
336        {
337          "Text": "prueba",
338          "Offset": 40900000,
339          "Duration": 1000000,
340          "Confidence": 0.9337882,
341          "Lexical": "prueba",
342          "ITN": "prueba",
343          "MaskedITN": "prueba",
344          "Display": "prueba",
345          "Words": ["prueba"]
346        },
347        {
348          "Text": "de",
349          "Offset": 41900000,
350          "Duration": 1000000,
351          "Confidence": 0.9337882,
352          "Lexical": "de",
353          "ITN": "de",
354          "MaskedITN": "de",
355          "Display": "de",
356          "Words": ["de"]
357        },
358        {
359          "Text": "una",
360          "Offset": 42900000,
361          "Duration": 1000000,
362          "Confidence": 0.9337882,
363          "Lexical": "una",
364          "ITN": "una",
365          "MaskedITN": "una",
366          "Display": "una",
367          "Words": ["una"]
368        },
369        {
370          "Text": "prueba",
371          "Offset": 43900000,
372          "Duration": 1000000,
373          "Confidence": 0.9337882,
374          "Lexical": "prueba",
375          "ITN": "prueba",
376          "MaskedITN": "prueba",
377          "Display": "prueba",
378          "Words": ["prueba"]
379        },
380        {
381          "Text": "de",
382          "Offset": 44900000,
383          "Duration": 1000000,
384          "Confidence": 0.9337882,
385          "Lexical": "de",
386          "ITN": "de",
387          "MaskedITN": "de",
388          "Display": "de",
389          "Words": ["de"]
390        },
391        {
392          "Text": "una",
393          "Offset": 45900000,
394          "Duration": 1000000,
395          "Confidence": 0.9337882,
396          "Lexical": "una",
397          "ITN": "una",
398          "MaskedITN": "una",
399          "Display": "una",
400          "Words": ["una"]
401        },
402        {
403          "Text": "prueba",
404          "Offset": 46900000,
405          "Duration": 1000000,
406          "Confidence": 0.9337882,
407          "Lexical": "prueba",
408          "ITN": "prueba",
409          "MaskedITN": "prueba",
410          "Display": "prueba",
411          "Words": ["prueba"]
412        },
413        {
414          "Text": "de",
415          "Offset": 47900000,
416          "Duration": 1000000,
417          "Confidence": 0.9337882,
418          "Lexical": "de",
419          "ITN": "de",
420          "MaskedITN": "de",
421          "Display": "de",
422          "Words": ["de"]
423        },
424        {
425          "Text": "una",
426          "Offset": 48900000,
427          "Duration": 1000000,
428          "Confidence": 0.9337882,
429          "Lexical": "una",
430          "ITN": "una",
431          "MaskedITN": "una",
432          "Display": "una",
433          "Words": ["una"]
434        },
435        {
436          "Text": "prueba",
437          "Offset": 49900000,
438          "Duration": 1000000,
439          "Confidence": 0.9337882,
440          "Lexical": "prueba",
441          "ITN": "prueba",
442          "MaskedITN": "prueba",
443          "Display": "prueba",
444          "Words": ["prueba"]
445        },
446        {
447          "Text": "de",
448          "Offset": 50900000,
449          "Duration": 1000000,
450          "Confidence": 0.9337882,
451          "Lexical": "de",
452          "ITN": "de",
453          "MaskedITN": "de",
454          "Display": "de",
455          "Words": ["de"]
456        },
457        {
458          "Text": "una",
459          "Offset": 51900000,
460          "Duration": 1000000,
461          "Confidence": 0.9337882,
462          "Lexical": "una",
463          "ITN": "una",
464          "MaskedITN": "una",
465          "Display": "una",
466          "Words": ["una"]
467        },
468        {
469          "Text": "prueba",
470          "Offset": 52900000,
471          "Duration": 1000000,
472          "Confidence": 0.9337882,
473          "Lexical": "prueba",
474          "ITN": "prueba",
475          "MaskedITN": "prueba",
476          "Display": "prueba",
477          "Words": ["prueba"]
478        },
479        {
480          "Text": "de",
481          "Offset": 53900000,
482          "Duration": 1000000,
483          "Confidence": 0.9337882,
484          "Lexical": "de",
485          "ITN": "de",
486          "MaskedITN": "de",
487          "Display": "de",
488          "Words": ["de"]
489        },
490        {
491          "Text": "una",
492          "Offset": 54900000,
493          "Duration": 1000000,
494          "Confidence": 0.9337882,
495          "Lexical": "una",
496          "ITN": "una",
497          "MaskedITN": "una",
498          "Display": "una",
499          "Words": ["una"]
500        },
501        {
502          "Text": "prueba",
503          "Offset": 55900000,
504          "Duration": 1000000,
505          "Confidence": 0.9337882,
506          "Lexical": "prueba",
507          "ITN": "prueba",
508          "MaskedITN": "prueba",
509          "Display": "prueba",
510          "Words": ["prueba"]
511        },
512        {
513          "Text": "de",
514          "Offset": 56900000,
515          "Duration": 1000000,
516          "Confidence": 0.9337882,
517          "Lexical": "de",
518          "ITN": "de",
519          "MaskedITN": "de",
520          "Display": "de",
521          "Words": ["de"]
522        },
523        {
524          "Text": "una",
525          "Offset": 57900000,
526          "Duration": 1000000,
527          "Confidence": 0.9337882,
528          "Lexical": "una",
529          "ITN": "una",
530          "MaskedITN": "una",
531          "Display": "una",
532          "Words": ["una"]
533        },
534        {
535          "Text": "prueba",
536          "Offset": 58900000,
537          "Duration": 1000000,
538          "Confidence": 0.9337882,
539          "Lexical": "prueba",
540          "ITN": "prueba",
541          "MaskedITN": "prueba",
542          "Display": "prueba",
543          "Words": ["prueba"]
544        },
545        {
546          "Text": "de",
547          "Offset": 59900000,
548          "Duration": 1000000,
549          "Confidence": 0.9337882,
550          "Lexical": "de",
551          "ITN": "de",
552          "MaskedITN": "de",
553          "Display": "de",
554          "Words": ["de"]
555        },
556        {
557          "Text": "una",
558          "Offset": 60900000,
559          "Duration": 1000000,
560          "Confidence": 0.9337882,
561          "Lexical": "una",
562          "ITN": "una",
563          "MaskedITN": "una",
564          "Display": "una",
565          "Words": ["una"]
566        },
567        {
568          "Text": "prueba",
569          "Offset": 61900000,
570          "Duration": 1000000,
571          "Confidence": 0.9337882,
572          "Lexical": "prueba",
573          "ITN": "prueba",
574          "MaskedITN": "prueba",
575          "Display": "prueba",
576          "Words": ["prueba"]
577        },
578        {
579          "Text": "de",
580          "Offset": 62900000,
581          "Duration": 1000000,
582          "Confidence": 0.9337882,
583          "Lexical": "de",
584          "ITN": "de",
585          "MaskedITN": "de",
586          "Display": "de",
587          "Words": ["de"]
588        },
589        {
590          "Text": "una",
591          "Offset": 63900000,
592          "Duration": 1000000,
593          "Confidence": 0.9337882,
594          "Lexical": "una",
595          "ITN": "una",
596          "MaskedITN": "una",
597          "Display": "una",
598          "Words": ["una"]
599        },
600        {
601          "Text": "prueba",
602          "Offset": 64900000,
603          "Duration": 1000000,
604          "Confidence": 0.9337882,
605          "Lexical": "prueba",
606          "ITN": "prueba",
607          "MaskedITN": "prueba",
608          "Display": "prueba",
609          "Words": ["prueba"]
610        },
611        {
612          "Text": "de",
613          "Offset": 65900000,
614          "Duration": 1000000,
615          "Confidence": 0.9337882,
616          "Lexical": "de",
617          "ITN": "de",
618          "MaskedITN": "de",
619          "Display": "de",
620          "Words": ["de"]
621        },
622        {
623          "Text": "una",
624          "Offset": 66900000,
625          "Duration": 1000000,
626          "Confidence": 0.9337882,
627          "Lexical": "una",
628          "ITN": "una",
629          "MaskedITN": "una",
630          "Display": "una",
631          "Words": ["una"]
632        },
633        {
634          "Text": "prueba",
635          "Offset": 67900000,
636          "Duration": 1000000,
637          "Confidence": 0.9337882,
638          "Lexical": "prueba",
639          "ITN": "prueba",
640          "MaskedITN": "prueba",
641          "Display": "prueba",
642          "Words": ["prueba"]
643        },
644        {
645          "Text": "de",
646          "Offset": 68900000,
647          "Duration": 1000000,
648          "Confidence": 0.9337882,
649          "Lexical": "de",
650          "ITN": "de",
651          "MaskedITN": "de",
652          "Display": "de",
653          "Words": ["de"]
654        },
655        {
656          "Text": "una",
657          "Offset": 69900000,
658          "Duration": 1000000,
659          "Confidence": 0.9337882,
660          "Lexical": "una",
661          "ITN": "una",
662          "MaskedITN": "una",
663          "Display": "una",
664          "Words": ["una"]
665        },
666        {
667          "Text": "prueba",
668          "Offset": 70900000,
669          "Duration": 1000000,
670          "Confidence": 0.9337882,
671          "Lexical": "prueba",
672          "ITN": "prueba",
673          "MaskedITN": "prueba",
674          "Display": "prueba",
675          "Words": ["prueba"]
676        },
677        {
678          "Text": "de",
679          "Offset": 71900000,
680          "Duration": 1000000,
681          "Confidence": 0.9337882,
682          "Lexical": "de",
683          "ITN": "de",
684          "MaskedITN": "de",
685          "Display": "de",
686          "Words": ["de"]
687        },
688        {
689          "Text": "una",
690          "Offset": 72900000,
691          "Duration": 1000000,
692          "Confidence": 0.9337882,
693          "Lexical": "una",
694          "ITN": "una",
695          "MaskedITN": "una",
696          "Display": "una",
697          "Words": ["una"]
698        },
699        {
700          "Text": "prueba",
701          "Offset": 73900000,
702          "Duration": 1000000,
703          "Confidence": 0.9337882,
704          "Lexical": "prueba",
705          "ITN": "prueba",
706          "MaskedITN": "prueba",
707          "Display": "prueba",
708          "Words": ["prueba"]
709        },
710        {
711          "Text": "de",
712          "Offset": 74900000,
713          "Duration": 1000000,
714          "Confidence": 0.9337882,
715          "Lexical": "de",
716          "ITN": "de",
717          "MaskedITN": "de",
718          "Display": "de",
719          "Words": ["de"]
720        },
721        {
722          "Text": "una",
723          "Offset": 75900000,
724          "Duration": 1000000,
725          "Confidence": 0.9337882,
726          "Lexical": "una",
727          "ITN": "una",
728          "MaskedITN": "una",
729          "Display": "una",
730          "Words": ["una"]
731        },
732        {
733          "Text": "prueba",
734          "Offset": 76900000,
735          "Duration": 1000000,
736          "Confidence": 0.9337882,
737          "Lexical": "prueba",
738          "ITN": "prueba",
739          "MaskedITN": "prueba",
740          "Display": "prueba",
741          "Words": ["prueba"]
742        },
743        {
744          "Text": "de",
745          "Offset": 77900000,
746          "Duration": 1000000,
747          "Confidence": 0.9337882,
748          "Lexical": "de",
749          "ITN": "de",
750          "MaskedITN": "de",
751          "Display": "de",
752          "Words": ["de"]
753        },
754        {
755          "Text": "una",
756          "Offset": 78900000,
757          "Duration": 1000000,
758          "Confidence": 0.9337882,
759          "Lexical": "una",
760          "ITN": "una",
761          "MaskedITN": "una",
762          "Display": "una",
763          "Words": ["una"]
764        },
765        {
766          "Text": "prueba",
767          "Offset": 79900000,
768          "Duration": 1000000,
769          "Confidence": 0.9337882,
770          "Lexical": "prueba",
771          "ITN": "prueba",
772          "MaskedITN": "prueba",
773          "Display": "prueba",
774          "Words": ["prueba"]
775        },
776        {
777          "Text": "de",
778          "Offset": 80900000,
779          "Duration": 1000000,
780          "Confidence": 0.9337882,
781          "Lexical": "de",
782          "ITN": "de",
783          "MaskedITN": "de",
784          "Display": "de",
785          "Words": ["de"]
786        },
787        {
788          "Text": "una",
789          "Offset": 81900000,
790          "Duration": 1000000,
791          "Confidence": 0.9337882,
792          "Lexical": "una",
793          "ITN": "una",
794          "MaskedITN": "una",
795          "Display": "una",
796          "Words": ["una"]
797        },
798        {
799          "Text": "prueba",
800          "Offset": 82900000,
801          "Duration": 1000000,
802          "Confidence": 0.9337882,
803          "Lexical": "prueba",
804          "ITN": "prueba",
805          "MaskedITN": "prueba",
806          "Display": "prueba",
807          "Words": ["prueba"]
808        },
809        {
810          "Text": "de",
811          "Offset": 83900000,
812          "Duration": 1000000,
813          "Confidence": 0.9337882,
814          "Lexical": "de",
815          "ITN": "de",
816          "MaskedITN": "de",
817          "Display": "de",
818          "Words": ["de"]
819        },
820        {
821          "Text": "una",
822          "Offset": 84900000,
823          "Duration": 1000000,
824          "Confidence": 0.9337882,
825          "Lexical": "una",
826          "ITN": "una",
827          "MaskedITN": "una",
828          "Display": "una",
829          "Words": ["una"]
830        },
831        {
832          "Text": "prueba",
833          "Offset": 85900000,
834          "Duration": 1000000,
835          "Confidence": 0.9337882,
836          "Lexical": "prueba",
837          "ITN": "prueba",
838          "MaskedITN": "prueba",
839          "Display": "prueba",
840          "Words": ["prueba"]
841        },
842        {
843          "Text": "de",
844          "Offset": 86900000,
845          "Duration": 1000000,
846          "Confidence": 0.9337882,
847          "Lexical": "de",
848          "ITN": "de",
849          "MaskedITN": "de",
850          "Display": "de",
851          "Words": ["de"]
852        },
853        {
854          "Text": "una",
855          "Offset": 87900000,
856          "Duration": 1000000,
857          "Confidence": 0.9337882,
858          "Lexical": "una",
859          "ITN": "una",
860          "MaskedITN": "una",
861          "Display": "una",
862          "Words": ["una"]
863        },
864        {
865          "Text": "prueba",
866          "Offset": 88900000,
867          "Duration": 1000000,
868          "Confidence": 0.9337882,
869          "Lexical": "prueba",
870          "ITN": "prueba",
871          "MaskedITN": "prueba",
872          "Display": "prueba",
873          "Words": ["prueba"]
874        },
875        {
876          "Text": "de",
877          "Offset": 89900000,
878          "Duration": 1000000,
879          "Confidence": 0.9337882,
880          "Lexical": "de",
881          "ITN": "de",
882          "MaskedITN": "de",
883          "Display": "de",
884          "Words": ["de"]
885        },
886        {
887          "Text": "una",
888          "Offset": 90900000,
889          "Duration": 1000000,
890          "Confidence": 0.9337882,
891          "Lexical": "una",
892          "ITN": "una",
893          "MaskedITN": "una",
894          "Display": "una",
895          "Words": ["una"]
896        },
897        {
898          "Text": "prueba",
899          "Offset": 91900000,
900          "Duration": 1000000,
901          "Confidence": 0.9337882,
902          "Lexical": "prueba",
903          "ITN": "prueba",
904          "MaskedITN": "prueba",
905          "Display": "prueba",
906          "Words": ["prueba"]
907        },
908        {
909          "Text": "de",
910          "Offset": 92900000,
911          "Duration": 1000000,
912          "Confidence": 0.9337882,
913          "Lexical": "de",
914          "ITN": "de",
915          "MaskedITN": "de",
916          "Display": "de",
917          "Words": ["de"]
918        },
919        {
920          "Text": "una",
921          "Offset": 93900000,
922          "Duration": 1000000,
923          "Confidence": 0.9337882,
924          "Lexical": "una",
925          "ITN": "una",
926          "MaskedITN": "una",
927          "Display": "una",
928          "Words": ["una"]
929        },
930        {
931          "Text": "prueba",
932         
```

5. Voz a texto (ASR) - Tiempo real.



Vamos recibiendo eventos de forma dinámica en formato JSON, hasta que terminamos la frase y se genera el final. Hemos pasado a RECOGNIZED.

5. Voz a texto (ASR) - Por lotes.



Con las transcripciones por lotes, se envían los datos de audio y, después, se recuperan los resultados de la transcripción de forma asincrónica. El servicio transcribe los datos de audio y almacena los resultados en un contenedor de almacenamiento. Después, podemos recuperar los resultados del contenedor de almacenamiento.

5. Voz a texto (ASR) - Por lotes.

Para hacer esta práctica necesitamos un modelo de servicios de voz que tenga la tarifa standard, en caso de no tener configurada esa crear un nuevo modelo, como hicimos en el TTS.

The screenshot displays the Azure portal interface for a resource named 'PruebaBatchASR' (Servicio de Voz). The left sidebar contains navigation options: 'Información general', 'Registro de actividad', 'Control de acceso (IAM)', 'Etiquetas', 'Diagnosticar y solucionar problemas', 'Administración de recursos', 'Claves y punto de conexión', 'Cifrado', and 'Precios del nivel de compromiso'. The main content area is divided into 'Información esencial' and 'Empezar'. The 'Información esencial' section lists details such as 'Grupo de recursos' (PruebaBatchASR), 'Estado' (Activo), 'Ubicación' (East US), 'Suscripción' (Azure for Students), 'Id. de suscripción' (9fd3c1cd-d793-4b97-8b31-084110699500), and 'Etiquetas' (Agregar etiquetas). The 'Empezar' section includes a 'Supervisión' tab and a prompt to 'Comience a usar el recurso en Speech Studio'. A table on the right provides additional configuration details: 'API Kind' (SpeechServices), 'Plan de tarifa' (Estándar), 'Extremo' (https://eastus.api.cognitive.microsoft.com/), 'Administrar claves' (Haga clic aquí para administrar claves), and 'Planes de compromiso' (Haga clic aquí para ver las opciones de precios del plan de compromiso).


Propiedad	Valor
API Kind	SpeechServices
Plan de tarifa	Estándar
Extremo	https://eastus.api.cognitive.microsoft.com/
Administrar claves	Haga clic aquí para administrar claves
Planes de compromiso	Haga clic aquí para ver las opciones de precios del plan de compromiso

5. Voz a texto (ASR) - Por lotes.

Seleccionamos
conversión de voz en
texto bath.

Conversión de voz en texto


Transcriba con rapidez y precisión en más de 100 idiomas y dialectos. Mejore la precisión de sus transcripciones creando un modelo de voz personalizado que pueda manejar la terminología específica del dominio, el ruido de fondo y los acentos. [Más información sobre la conversión de voz en texto](#)



Voz a texto en tiempo real

Pruebe de forma rápida las capacidades de transcripción en vivo en su propio audio sin escribir ningún código.


[Probar Voz a texto en tiempo real](#)



Modelo de susurro en el servicio Azure OpenAI

Pruebe rápidamente las funcionalidades de transcripción en directo en su propio audio con su recurso de Azure OpenAI y use avisos para mejorar la calidad de las transcripciones.


[Probar Modelo de susurro en el servicio Azure OpenAI](#)



Conversión de voz en texto de Batch

Pruebe rápidamente las capacidades de transcripción por lotes para transcribir una gran cantidad de audio almacenado y reciba resultados de forma asincrónica mediante los modelos Azure Speech o OpenAI Whisper.


[Probar Conversión de voz en texto de Batch](#)



Habla personalizada

Agregue sus propios datos y adapte a estilos de habla específicos, vocabulario y mucho más con un modelo personalizado de conversión de voz en texto.


[Iniciar un proyecto de habla personalizada](#)



Evaluación de pronunciación con conversión de voz en texto

Obtenga información instantánea sobre la precisión y la fluidez de la pronunciación leyendo un guión en voz alta.

[Probar Evaluación de la pronunciación](#)



Traducción de voz

Traduzca la voz a otros idiomas de su elección con baja latencia.

[Probar Traducción de voz](#)

5. Voz a texto (ASR) - Por lotes.

Idioma: Español (España)

Modelo: Whisper Large V2

Cargar varios archivos de audio, si no teneis podeis utilizar la grabadora de Windows para generar uno.

[Ver documentación](#) [Ver código de muestra](#) [Usar la CLI de Voz](#)

Probarlo

Idioma de audio: Idioma de audio Modelo Modelo que se va a comparar (opcional)

Idioma de audio Modelo Modelo que se va a comparar (opcional)

Idioma de audio Modelo Modelo que se va a comparar (opcional)

[Mostrar opciones avanzadas](#)

Usar con el recurso

☒ Reconozco que esta aplicación usa el recurso [Estenombretienequeserunico2](#) y que implicará el uso de mi cuenta. [Elige otro recurso](#)

Cargar archivos de audio

Archivo local Direcciones URL web

Arrastrar y colocar archivos

[Examinar archivos...](#)

Trabajos por lotes

Trabajos por lotes	20240311_114253
<input checked="" type="checkbox"/> 20240311_114253	<input checked="" type="checkbox"/> Grabación (2).m4a
	<input checked="" type="checkbox"/> Grabación (3).m4a
	<input checked="" type="checkbox"/> Grabación.m4a

[Borrar todo](#)

5. Voz a texto (ASR) - Por lotes.

Veréis el resultado de la prueba, tanto en formato Texto como JSON.

Resultados de pruebas

Modelo: 20240228 Whisper Large V2

Lenguaje: Español (España)

Diarización de altavoz: No

Modo de filtro de palabras soeces: Masked

Modo de puntuación: DictatedAndAutomatic

Marca de tiempo de nivel de Word: Sí (mostrar); No (éxico)

00:05

00:06

Texto

JSON

Hola, esto es la prueba 2 de la clase de modelos, prueba 2 de la clase de modelos.

6. Servicios de traducción



Traductor es un servicio de traducción automática neural basado en la nube que forma parte de la familia de servicios de Azure AI y que puede utilizarse con cualquier sistema operativo.

6. Servicios de traducción

Característica	Descripción	Opciones de desarrollo
Traducción de texto	Ejecute la traducción de texto en tiempo real entre los idiomas de origen y destino admitidos. Cree un diccionario dinámico y descubra cómo evitar traducciones mediante la API de Translator.	<ul style="list-style-type: none">• API REST• Contenedor de traducción de texto
Traducción asincrónica de documentos por lotes	Traduzca archivos complejos y por lotes sin perder la estructura y el formato de los documentos originales. Cree un glosario para usarlo con la traducción de documentos. El proceso de traducción por lotes requiere una cuenta de Azure Blob Storage con contenedores para los documentos de origen y traducidos.	<ul style="list-style-type: none">• API REST• SDK de biblioteca de cliente
Traducción sincrónica de documentos	Traduzca un solo archivo de documento o con un archivo de glosario al tiempo que conserva la estructura y el formato del documento original. El proceso de traducción de archivos no requiere una cuenta de Azure Blob Storage. La respuesta final contiene el documento traducido y se devuelve directamente al cliente que realiza la llamada.	REST API
Custom Translator	Cree modelos personalizados para traducir terminología y estilo específicos del sector y del dominio. Cree un diccionario (oración o frase) para traducciones personalizadas.	<ul style="list-style-type: none">• Portal de Traductor personalizado 

6. Servicios de traducción

Ahora vamos a probarlo

Conversión de voz en texto

Transcriba con rapidez y precisión en más de 100 idiomas y dialectos. Mejore la precisión de sus transcripciones creando un modelo de voz personalizado que pueda manejar la terminología específica del dominio, el ruido de fondo y los acentos. [Más información sobre la conversión de voz en texto](#)



Voz a texto en tiempo real

Pruebe de forma rápida las capacidades de transcripción en vivo en su propio audio sin escribir ningún código.

[Probar Voz a texto en tiempo real](#)



Modelo de susurro en el servicio Azure OpenAI

Pruebe rápidamente las funcionalidades de transcripción en directo en su propio audio con su recurso de Azure OpenAI y use avisos para mejorar la calidad de las transcripciones.

[Probar Modelo de susurro en el servicio Azure OpenAI](#)



Conversión de voz en texto de Batch

Pruebe rápidamente las capacidades de transcripción por lotes para transcribir una gran cantidad de audio almacenado y reciba resultados de forma asincrónica mediante los modelos Azure Speech o OpenAI Whisper.

[Probar Conversión de voz en texto de Batch](#)



Habla personalizada

Agregue sus propios datos y adapte a estilos de habla específicos, vocabulario y mucho más con un modelo personalizado de conversión de voz en texto.

[Iniciar un proyecto de habla personalizada](#)



Evaluación de pronunciación con conversión de voz en texto

Obtenga información instantánea sobre la precisión y la fluidez de la pronunciación leyendo un guión en voz alta.

[Probar Evaluación de la pronunciación](#)



Traducción de voz

Traduzca la voz a otros idiomas de su elección con baja latencia.

[Probar Traducción de voz](#)

6. Servicios de traducción

Subimos un archivo y lo probamos

☒ Reconozco que esta aplicación usa el recurso Estenombretienequeserunico2 y que implicará el uso de mi cuenta. [Elige otro recurso](#)

Elegir un idioma hablado

Español (España)

Elegir un idioma destino

Inglés

Leer traducción ☐ Nombre de voz: Jenny (en-US-JennyNeural)

Elegir los archivos de audio

Arrastrar y colocar archivos de audio aquí o
[Examinar archivos...](#)

O grabar audio con un micrófono

Archivos de audio

Grabación.m4a

[Borrar todo](#)

Resultados de pruebas

Nombre de archivo: Grabación.m4a

Idioma de destino: Inglés

Lenguaje: Español (España)

Nombre de voz: Jenny (en-US-JennyNeural)

00:00 00:04s

Traducida texto Original texto JSON

Hello, this is a test of the kind of models.

6. Servicios de traducción

En el JSON podemos ver como hace primero la conversión y después la traducción.



6. Servicios de traducción



Ahora vamos a hacer el Laboratorio 3 Transcribir, traducir y reproducir audios

Nos descargamos el Google Colab de Aules.