

# Reconocimiento automático del habla

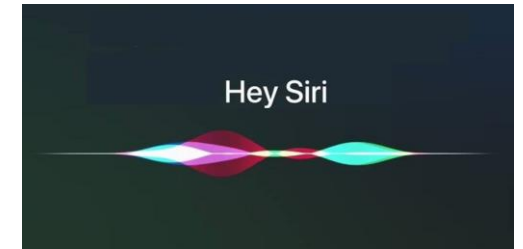
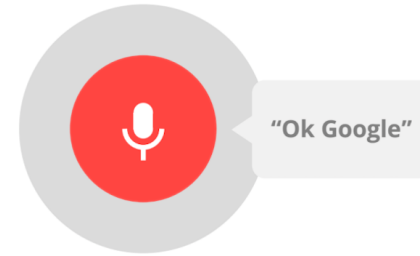
---

CURSO GRUPO BANCOLOMBIA

Universidad Nacional de Colombia

# Reconocimiento automático del habla

- El reconocimiento automático del habla busca permitir la comunicación entre computadores y seres humanos, de modo tal que la máquina sea capaz de entender el sonido generado por el humano.
- Grandes retos surgen en este campo ya que existen muchas variables que afectan la capacidad de entendimiento, como el idioma, el acento, la fonética, la acústica, el ruido, entre otros.
- En muchas ocasiones se trabaja sobre un subconjunto de toda la gramática disponible, con el fin de realizar modelos más precisos en un campo particular. Por ejemplo, un sistema puede trabajar mejor si el humano responde a preguntas guiadas y cortas en un contexto particular.

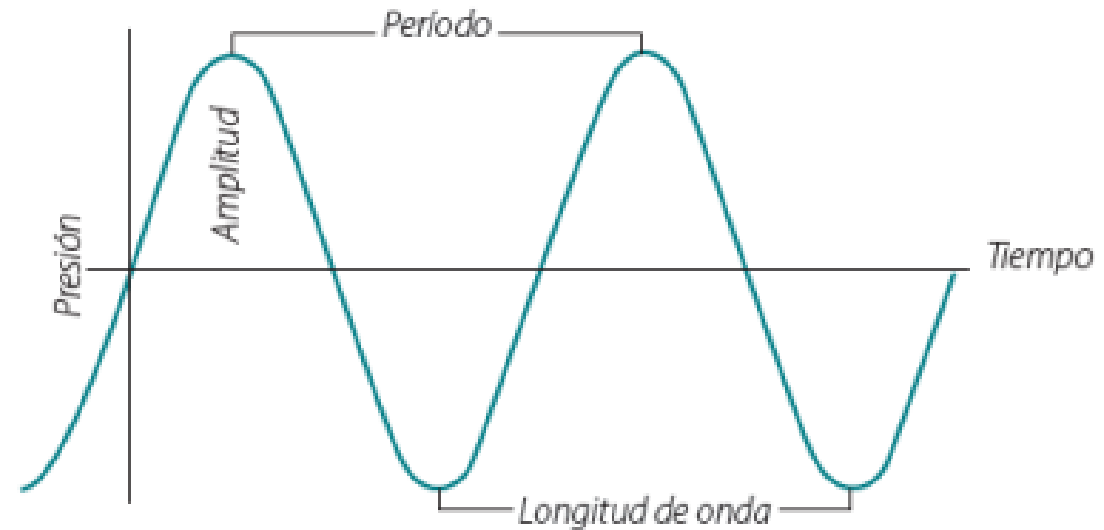


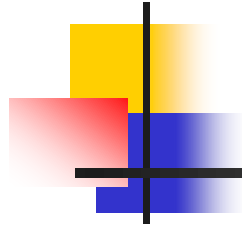
"Alexa, play morning playlist."



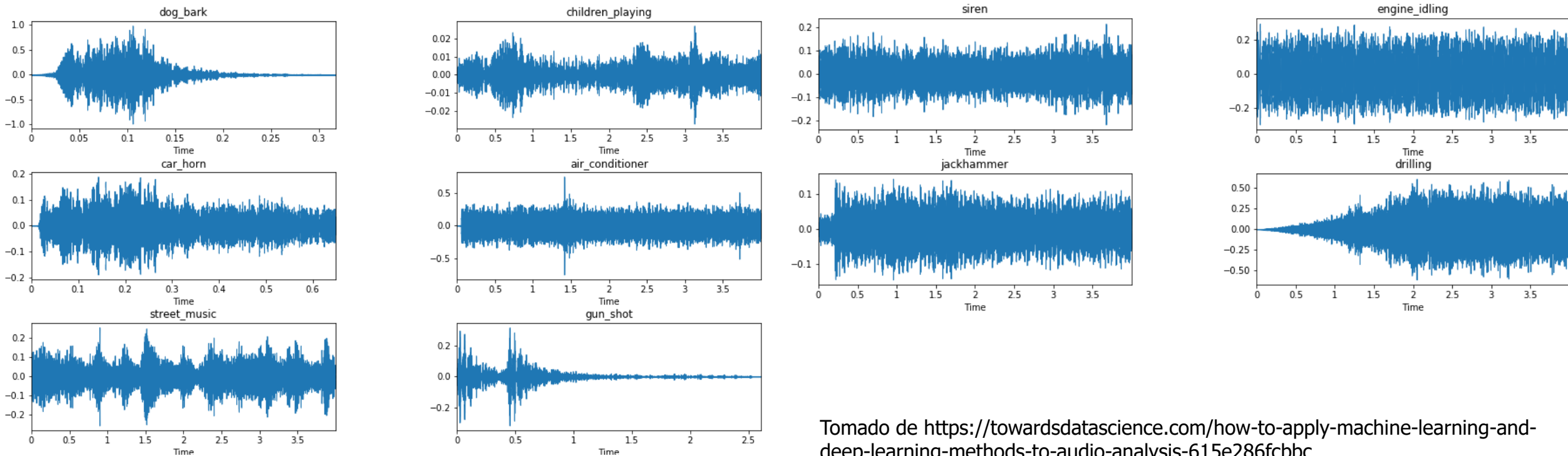
# Reconocimiento automático del habla

- El reconocimiento del habla puede ser visto como un problema de clasificación sobre patrones. Cada sonido genera ondas que se pueden analizar en intervalos cortos, en busca de patrones.
- Se debe analizar el sonido en segmentos de corta duración para reconocer patrones y luego asignarlos a elementos del lenguaje en un proceso de clasificación.
- Se pueden encontrar patrones y elementos similares en ondas que representan el mismo sonido, lo cual permite utilizar modelos de machine learning para esta tarea.





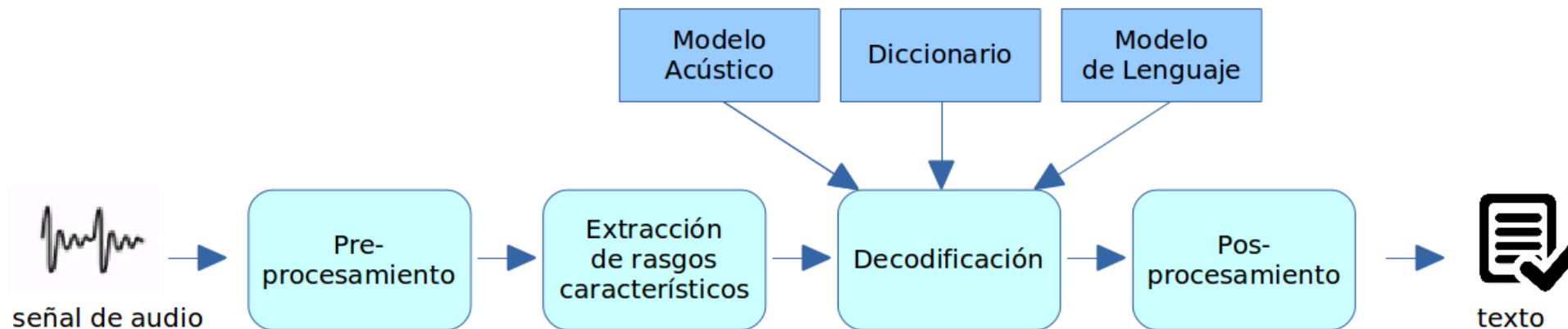
# Reconocimiento automático del habla



Tomado de <https://towardsdatascience.com/how-to-apply-machine-learning-and-deep-learning-methods-to-audio-analysis-615e286fcbbc>

# Reconocimiento automático del habla

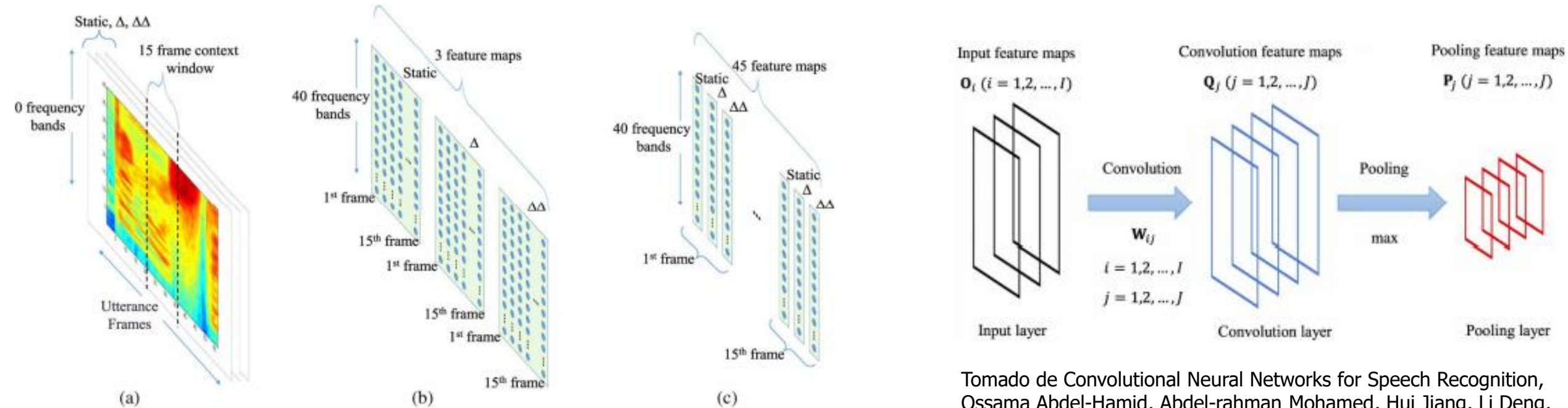
El proceso consta de varias etapas: Una fase de pre-procesamiento donde se realizan conversiones a ciertas frecuencias y se aplican filtros para reducir el ruido; una fase de extracción de características donde se buscan patrones; decodificación en la que se buscan los elementos del lenguaje más probables de corresponder a la señal; y un post-procesamiento donde se pueden aplicar herramientas propias de cada contexto.



Tomado de <https://medium.com/soldai/c%C3%B3mo-funciona-el-reconocimiento-autom%C3%A1tico-del-habla-eb038ecfe72e>

# Deep learning

- Las redes convolucionales son ampliamente utilizadas para clasificación de audios gracias a su capacidad para extraer características. Su funcionamiento es muy similar al reconocimiento de imágenes.



Tomado de Convolutional Neural Networks for Speech Recognition, Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu



¡A codificar!

