

# Reconocimiento de Personas Por Medio Voz, Aplicado a un Robot Social

Aplicación de una técnica de Machine Learning para el  
reconocimiento de personas a través de la voz.

Alvaro Araujo

10 de abril de 2018

La voz, es una característica distintiva de cada ser humano, aunque existen métodos biometricos ampliamente utilizados para reconocer de una forma segura y robusta a cada individuo, como es el caso de las huellas dactilares, la voz contiene un contexto social, es una de las formas en la cual las personas reconocen a sus semejantes. ¿Es común que una persona reconozca a otra observando sus huellas dactilares? Por supuesto que no. En las personas, normalmente la forma de reconocer a otras personas es a través de los sentidos del oído y la vista.

# El Oído Humano

El oído es un órgano maravilloso, que permite a las personas no solo captar la información que se les trasmite a través de la voz, si no que posee la agudeza para capturar la información correspondiente a muchos sonidos en un amplio rango de frecuencias, gracias a el, las personas son capaces de disfrutar la música y reconocer a otras personas con la características presentes en su voz.

## **Aplicación del Reconocimiento de Voz de un Hablante Mediante una Red Neuronal Artificial Backpropagation y Coeficientes LPC sobre un Canal Telefónico.**

- Cruz, L. and Acevedo, M.
- Programación Matemática y Software.
- Vol. 1, No 1, 2009, ISSN: 2007-3283
- Sitio web de la revista

## **Sistema de reconocimiento de voz mediante wavelets, predicción lineal y redes backpropagation.**

- San Juan, E. ; Jamett, M. ; Kaschel, H. and Sánchez L.
- Ingeniare. Revista chilena de ingeniería.
- Vol. 24 No 1, 2016, pp. 8-17
- Sitio web de la revista

# Justificación

En el campo de la robótica social, es de gran importancia estudiar formas de crear una mejor interacción entre los humanos y los robots; Se espera, que este componente de software tenga un efecto positivo sobre los usuarios del robot, partiendo de la idea de que un robot capaz de reconocer al usuario en cuestión, con solo escuchar su voz pueda crear un lazo social con el mismo.

# Propuesta

Se propone construir un componente de software para un robot social, que sea capaz de reconocer a una persona por medio de su voz, además de diferenciar la voz de este de otros posibles usuarios. Se propone una implementación utilizando Redes Neuronales.

El componente propuesto en esta instancia busca poder reconocer a una persona analizando solo un saludo, la frase a utilizar será “Hola Robot!” a la cual el componente deberá responder “Hola [usuario]!”.



# Wavelets

La transformada de ondícula es un tipo especial de transformada matemática que representa una señal en términos de versiones trasladadas y dilatadas de una onda finita (denominada ondula madre). La teoría de ondículas está relacionada con campos muy variados. Todas las transformaciones de ondículas pueden ser consideradas formas de representación en tiempo-frecuencia y, por tanto, están relacionadas con el análisis armónico.

# Coeficientes LPC (Codificación Predictiva Lineal)

La codificación predictiva lineal es un tipo de codificador ampliamente utilizado en audio digital. En sistemas de procesamiento de voz, se usa partiendo de la idea de que la voz puede modelarse como una combinación lineal de  $p$  muestras anteriores más una señal de error.

# Coeficientes Ceptrales, Escala de Frecuencias Mel

Los coeficientes ceptrales en la escala de frecuencias de Mel, son coeficientes para la representación del habla basados en la percepción auditiva humana. Estos surgen de la necesidad, en el área del reconocimiento de audio automático, de extraer características de una señal de audio que sean adecuadas para la identificación de contenido relevante, así como obviar todas aquellas que posean información poco valiosa como el ruido de fondo, emociones, volumen, tono, entre otras.

# Redes Neuronales Backpropagation

En este tipo de redes neuronales para su entrenamiento se utiliza un método de cálculo del gradiente con propagación hacia atrás de errores.

Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo las neuronas de la capa oculta solo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total.

# Propuesta de implementación

- Implementar un preprocesador de audio que cumpla dos funciones, quitar el ruido (10 % de la amplitud máxima) y normalizar la señal en el rango  $[1,-1]$ .
- Armar un conjunto de prueba inicial pequeño con la estructura antes mencionada, la frase “Hola Robot!” etiquetada con el nombre del usuario que la pronuncio.
- Implementar una Red Neuronal Backpropagation.
- Extraer la información numérica de la muestra de audio preprocesada y entrenar a la red neuronal con estos datos.

# Propuesta de implementación

- Validar el modelo a una aproximación del objetivo.
- Estudiar la factibilidad de extraer los coeficientes ceptrales o los coeficientes de codificación de predicción lineal y entrenar la red neuronal con esta información.
- Construir una base de datos con una cantidad de muestras adecuada y validar los modelos y compararlos (en caso de haber implementado mas de uno). Finalmente presentar los resultados.