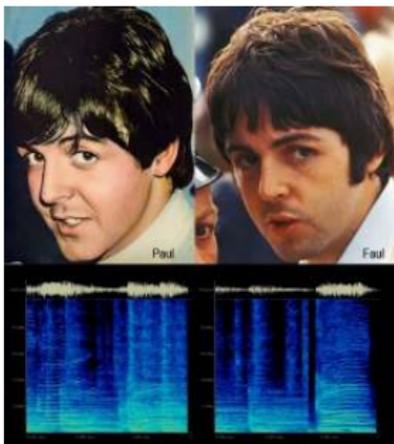


¿Paul McCartney es un impostor?

Introducción al Reconocimiento Forense de Hablantes



Miguel Martínez Soler
Laboratorio de Investigaciones Sensoriales (LIS)
INIGEM, CONICET-UBA, Hospital de Clínicas

Nuestros personajes en 1962



- Paul McCartney (primero desde la izquierda)
- 20 años
- Durante este año grabaría con The Beatles su primer disco



LAWRENCE G. KERSTA

B.S., M.S., Columbia University; former President of Voiceprint Laboratory and research scientist for Bell Telephone Laboratories; founder of Voiceprint Identification Systems; author and lecturer.

VOICEPRINT IDENTIFICATION
By Lawrence G. Kersta

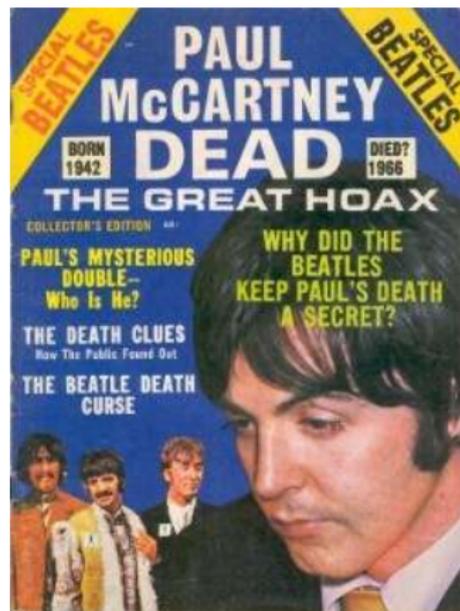
- Lawrence Kersta
- 55 años
- Durante este año publicaría su primer paper en Nature (otro vendría en 1968)

1960s - Beatlemania

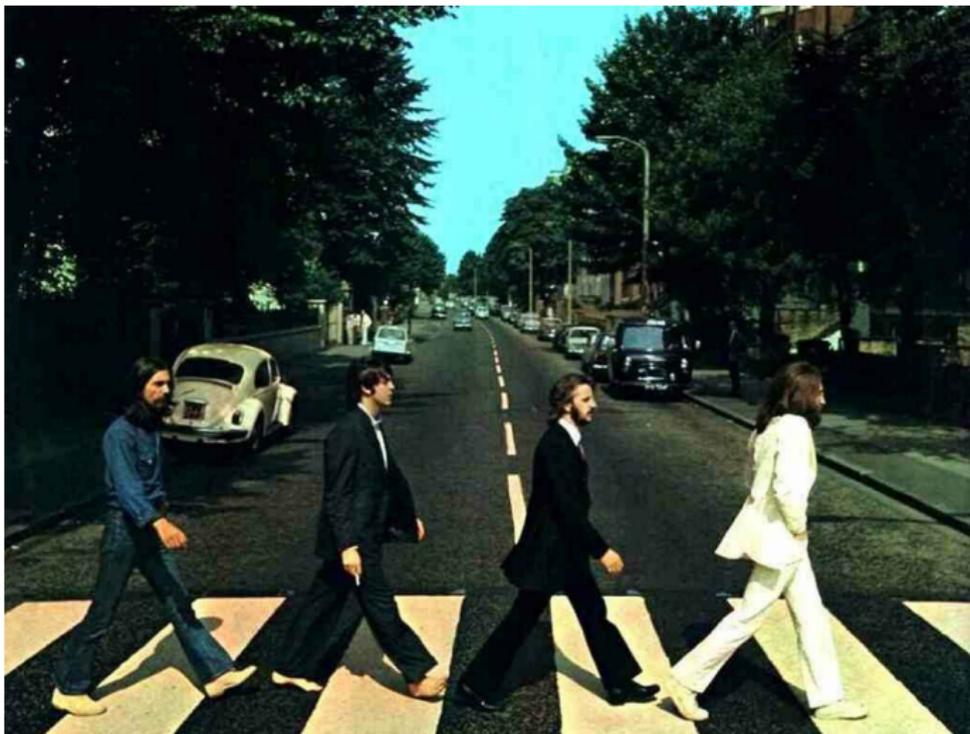


Vamos llegando al punto...

- 1966: La Beatlemanía llega a Japón
- Enero de 1967: Rumor de la muerte de Paul
- 1969: Un periódico estudiantil de la U. de Drake publica el artículo "Is Beatle Paul McCartney Dead?"
- El rumor se extendió muy pronto por los medios



La evidencia



44 años después...

Tengo que dar una charla de borrachos sobre reconocimiento forense de hablantes y necesito un hilo conductor... busquemos.



Ahora COMPAREMOS Ambos ESPECTROS

Paul McCartney Vs Faul McCartney, Forensic Voice Comparison

liteneutrino · 70 videos

Suscribirse 520

13.626

45 14

The image shows a YouTube video player. The video content is a spectrogram titled "Ahora COMPAREMOS Ambos ESPECTROS" (Now let's compare both spectra). It displays two side-by-side spectrograms of a voice recording, likely comparing Paul McCartney and a look-alike (Faul McCartney). The spectrograms show frequency components over time, with a red progress bar at the bottom indicating the video is at 02:09 / 15:01. Below the video, the title "Paul McCartney Vs Faul McCartney, Forensic Voice Comparison" is visible, along with the channel name "liteneutrino" (70 videos), a "Suscribirse" (Subscribe) button with 520 subscribers, and engagement metrics: 13,626 views, 45 likes, and 14 comments.

Paul vs. Faul vs. Terry

- Sobre Paul en 1962 cantando P.S. I love you: “Es una voz GRUESA con amplitud hacia TONOS GRAVES y a la vez AGUDOS” [Play Sound](#)
- Sobre Faul en 1966 cantando Hello Goodbye: “Definitivamente una voz parecida pero con un tono diferente y sobre todo AGUDA” [Play Sound](#)
- Sobre Terry en 1967 cantando The Fool on the Hill: “Sin necesidad de apoyarnos en un espectro, es FACIL notar que Terry tiene una voz Aguda cuando así lo desea y GRUESA mucho más parecida a la de Paul...â” [Play Sound](#)

Paul vs. Faul vs. Terry

- ¿Quién cree que Paul, Faul y Terri son la misma persona?
- ¿Quién cree que Paul, Faul y Terri son personas DISTINTAS?
- ¿Quién cree que Faul y Terri son la misma persona DISTINTA DE PAUL?
- ¿Alguien cree en las demás opciones?
($Paul == Faul \neq Terri \parallel Paul \neq Faul == Terri$)

¿Se acuerdan de las "voiceprints" ?

- Mientras Kersta insistía con su método en 1973
- Oscar Tosi de la Universidad del Estado de Michigan reportó un resultado en 1972 del 0.5 % bajo estas condiciones:
 - 9 palabras aisladas
 - identificación cerrada
 - habla contemporánea
- Esto no pasa nunca en las situaciones reales. Tosi repitió los experimentos en estas condiciones:
 - palabras no aisladas
 - identificación abierta
 - habla NO contemporánea
- El error subió al 18 %. Tosi concluyó que el error sería insignificante si los peritos toman la decisión solamente cuando están completamente seguros (2/3 de los errores habían sido etiquetados como "uncertain").

Una visión crítica de las “voiceprints”

- La comunidad científica no ve con buenos ojos a la identificación por impresiones vocales:
 - No es un procedimiento objetivo
 - La performance está demasiado influida por las condiciones
 - Las evaluaciones en la literatura son ambiguas
- Lo que pasa es que la señal no es función del hablante sólomente, sino también de:
 - El mensaje
 - El estado emocional
 - El interlocutor
 - El equipo de grabación
 - El estado de salud
 - La edad
 - El ambiente
 - Un largo y siempre sorprendente etc.

¿Cómo impacta esto en nuestro intento de identificar a Paul?

- Las grabaciones de Paul que escuchamos tienen instrumentos sonando a un volumen considerable
- Los cantantes impostan la voz de formas diferentes para conseguir timbres distintos
- Durante el proceso de grabación hay un ingeniero de sonido que filtra, mezcla y comprime la señal
- Esto se agrega a todo lo de la diapositiva anterior, que sigue siendo válido

Variabilidad

Se identifican dos tipos de variabilidad para el problema del reconocimiento:

- **Intra-hablante:** Las voz o la grabación de la voz de una misma persona cambia según las circunstancias.
- **Inter-hablante:** Las voces de personas distintas presentan patrones diferentes en la voz

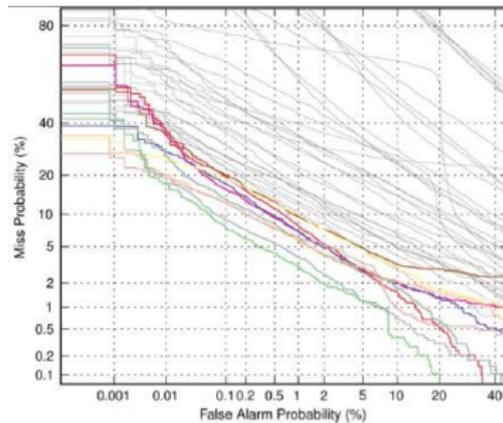


1980s en adelante

- Otras alternativas:
 - Redes Neuronales (por ej. Oglesby, 1990)
 - Redes Neuronales Auto-asociativas + Factor Analysis (Garimella y Hermansky, 2012)
 - Redes Neuronales de aprendizaje profundo con front-end UBM (Senoussaoui et al. 2012)
 - Redes Neuronales de aprendizaje profundo end-to-end ???

Tipos de Error

- Falsa alarma (FA): cuando decimos que ambas voces pertenecen a la misma persona, cuando en realidad no es así
- Pérdida (Miss): cuando decimos que las voces pertenecen a personas diferentes, cuando en realidad son de la misma persona



¿Lo podemos usar en la corte?

- Año 1993: U.S. Supreme Court, "Daubert v. Merrell Dow Pharmaceuticals 509 U.S. 579," 1993.
- La Corte Suprema de Estados Unidos emite el primero de una serie de fallos sobre la admisibilidad de evidencia científica.
- Antes de esto, se seguía el estándar Frye: La opinión de un experto es admitida solamente si su método es generalmente aceptado por la comunidad científica.
- Según el estándar Daubert el juez debe admitir una técnica si:
 - Ha sido o puede ser testada
 - Ha pasado por un proceso de revisión por pares o publicación
 - Hay estándares controlando su uso
 - Es generalmente aceptada en la comunidad científica
 - Tiene un grado de error conocido y bajo
 - Debe ayudar al juez de los hechos a comprender la evidencia o determinar un hecho que se trata en el caso

¿Lo podemos usar en la corte?

En otras palabras, se pide

- **Transparencia:** que la técnica sea conocida, aceptada y reproducible
- **Testeabilidad:** que hayan protocolos de testeo y bases de datos estándar que permitan estimar el grado de error
- **Comprensibilidad:** no sirve de nada un reporte forense si sólo lo entiende el experto

¿Se consideran oyentes ingenuos?

En otras palabras, se pide

- NIST Human Assisted Speaker Recognition
- Es una evaluación de sistemas de reconocimiento de hablantes que pueden incluir participación humana en el proceso
- Shen W., Campbell J., Straub D., and Schwartz R..
Assessing the speaker recognition performance of naive listeners using mechanical turk. ICASSP, page 5916-5919. IEEE, (2011)
- La gente del MITLL usó Mechanical Turk para obtener respuestas humanas

Veamos que tan ingenuos somos...

Test

Veamos que tan ingenuos somos...

Resultados

Say live and let die!

