



**INSTITUTO POLITÉCNICO NACIONAL  
ESCUELA SUPERIOR DE CÓMPUTO  
COMUNICACIÓN ORAL Y ESCRITA**



**Algoritmo para la detección de errores de fonética de una persona hispanohablante al hablar inglés**

Pedro Emilio Vélez Lozano, pvelez @alumno.ipn.mx

Escuela Superior de Cómputo

**Resumen**

A lo largo del artículo se expondrá la investigación sobre la detección de errores fonéticos en el habla, el estudio del inglés y el español para detectarlos y cómo es que un algoritmo puede procesar las palabras. Fue hecho el proyecto para la clase de Comunicación Oral y Escrita de la Escuela Superior de Cómputo con la ayuda del departamento de investigación de la universidad.

Los resultados principales indican que clasificar las palabras en fonemas de personas hablando inglés es algo posible y que se implementará. La trascendencia del artículo se espera sea alta para continuar con su investigación y perfección.

**Abstract**

Throughout the article the investigation about the detection of phonetic errors will be exposed, the analysis of using English as our main criteria and how an algorithm can process spoken words. This project was made for the Oral and Written Communication of the Superior School of Computer Sciences with the help of the department of investigation of the university.

The principal results show that classifying words in phonemes of people speaking English is something possible and that will be implemented. The

significance of the article is expected to be highly enough for us to keep investigating and making the system better.

## **Introducción**

El reto para las personas al hablar un segundo idioma es enorme en todo el mundo. La cantidad de personas que quiere hablar un segundo o tercer idioma se incrementa conforme pasan los años. Para esto, un sistema que pueda detectar los errores que tiene una persona al pronunciar las palabras en el idioma inglés es necesario para mejorar el habla de las personas hispanohablantes.

Este sistema es de bastante utilidad y se puede usar como método de enseñanza para todas las personas.

## **Desarrollo**

### **I. Intervenciones en la investigación**

Tener conocimiento computacional y matemático fue indispensable para elaborar la investigación.

Tuvimos que aprender inteligencia artificial, más específico lenguaje natural para saber qué herramientas podemos usar para facilitarnos el trabajo de crear el algoritmo.

Usamos análisis vectorial para ver los vectores, así como los campos ortogonales y la desigualdad de Cauchy-Swacz

El uso del análisis de sumas y series para generar un número de compatibilidad.

Lingüística sobre todo para aprender teorías de aprendizaje de lenguaje y los errores que tiene al hablar otro idioma, apartado que se expone a continuación.

### **II. Errores al hablar otro idioma**

Los errores al hablar otro idioma son bastante comunes, tanto que incluso pasa entre las personas para las que están hablando en su idioma natal.

### III. El idioma inglés y el español

Elaboramos la investigación en inglés porque es el idioma que más habla la gente y es el idioma que más aprenden los hispanohablantes. Aparte de que la información necesaria para elaborar la investigación casi en su totalidad está en inglés y los corpus mandados fueron comparados de personas hablando inglés como su primer idioma.

Esto es bastante importante de señalar, porque la clasificación fonética varía bastante entre idiomas, por ejemplo, el francés tiene los fonemas más “pegados” y es más difícil poder deducir con vectores qué es lo que se está diciendo, por lo que el margen de error es altísimo.

El idioma inglés es un idioma relativamente fácil de aprender e identificar, de representar y hablar. Este aspecto es indispensable para la dificultad del algoritmo.

### IV. Algoritmo para la detección de errores

Se presentan varias técnicas usadas en la investigación para la clasificación de información y se exponen en qué consiste cada una

- PLSA

En este trabajo, se presenta una técnica estadística de aprendizaje no supervisado conocida como análisis semántico probabilístico latente (PLSA, por sus siglas en inglés), cuya idea consiste en mapear vectores de conteo de alta dimensión a una representación dimensional inferior.

- Redes neuronales profundas

Por ello, se propone un modelo fonémico acústico-graphémico (AGPM) utilizando redes neuronales profundas (DNN por sus siglas en inglés) de distribución múltiple, cuyas características de entrada incluyen características acústicas, así como los grafemas correspondientes y transcripciones canónicas (codificadas como vectores binarios).

Para la implementación de este método se utilizan los conjuntos de datos del TIMIT y CUCHLOE (estudiantes chinos universitarios chinos de ENGÉISH)

Además, se utilizan modelos tales como modelo acústico de nivel estatal (s-AM), un modelo de transición de estado (STM), un modelo fonémico acústico-graphémico (AGPM), un modelo de fonémico acústico (APM), reconocimiento de teléfono gratuito para L2 habla inglesa (BASELINE).

- VMS

Para alcanzar este objetivo, el sistema CALL tiene que ser entrenado con corpus orientados a un idioma en particular (dada la complejidad de crear un sistema que abarque todos los idiomas), en este caso, el idioma practicado por los vietnamitas.

Comparación de los métodos GOP, Weilget, LDA-APF y LDA-MFCC

- LDA

Para proveer de retroalimentación a los sistemas de aprendizaje de idiomas asistido por computadora, se requiere que los errores de pronunciación se detecten automáticamente. Por ello, se hacer una comparación entre cuatro clasificadores que pueden ser usados para este propósito: dos clasificadores acústico-fonético (uno de los cuales emplea análisis discriminante lineal (LDA, por sus siglas en inglés)), un clasificador con base en coeficientes cepstrales en combinación con LDA, y otro con base en medidas de confianza (conocido como puntuación de buena calidad de la pronunciación (GOP, por sus siglas en inglés)).

Para la obtención del nivel de exactitud que produce cada uno de estos métodos se utilizaron el discurso nativo de la base de datos polyphone, y discurso no nativo del corpus DL2N1 (holandés como segundo idioma, Nimega corpus 1).

A partir de un corpus de unos audios de personas hablando inglés como primer idioma y unos audios de personas hispanohablantes hablando inglés, introdujimos los audios en Python y los comparamos usando los métodos explicados.

## **Recomendaciones**

El trabajo pendiente por realizar para concluir con nuestro proyecto y nuestra propuesta de solución es crear formalmente el sistema en algún lenguaje de programación en donde implementemos de manera más dinámica el algoritmo que acabamos de crear.

Crear un sistema va a ser la parte visual de nuestro proyecto de investigación, en el que necesitamos conocimientos para crear aplicaciones de computadora, crear páginas web, una interfaz para usuario, etcétera.

## **Resultados**

Aplicamos los métodos de clasificación a nuestra investigación comparando diversos factores como:

1. Facilidad para entender la representación de los vectores
2. Exactitud
3. Velocidad para clasificar los datos

El uso de SVM es el más rápido para clasificar vectores y su representación en forma de números y proximidad es más exacto y fácil de entender que los otros métodos aunque es un poco más lento que otros algoritmos.

## **Conclusiones**

Gracias a los avances de aprendizaje, la representación de la fonética mediante vectores es exitosa y el uso de SVM es necesario para detectar los errores de pronunciación de una persona hispanohablante al hablar inglés.

Este es un paso agigantado para nosotros como sociedad, todo lo que se puede hacer con procesamiento de lenguaje natural es increíble y aporta bastante para la humanidad.

## Referencias

[1] Bobadilla, A. G. (1996). Análisis contrastivo y análisis de errores en el aula del español como lengua extranjera. In *Tendencias actuales en la enseñanza del español como lengua extranjera I: actas del quinto Congreso Internacional de ASELE: Santander, 29, 30 de septiembre y 1 de octubre de 1994* (pp. 101-110). Asociación para la Enseñanza del Español como Lengua Extranjera.

[2] Hofmann, T. (2001). *Unsupervised learning by probabilistic latent semantic analysis*. *Machine learning*, 42(1-2), 177-196. doi: <https://doi.org/10.1023/A:1007617005950>

[3] Li, K., Qian, X., & Meng, H. (2017). *Mispronunciation detection and diagnosis in l2 English speech using multi-distribution deep neural networks*. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(1), 193-207. doi: 10.1109/TASLP.2016.2621675

[4] Dang, T. D., & Thi, K. G. D. (2017). *Automatic detection of common mispronunciations of Vietnamese speakers of English using SVMs*. In *2017 International Conference on System Science and Engineering (ICSSE)* (pp. 231-234). IEEE. doi: [10.1109/ICSSE.2017.8030871](https://doi.org/10.1109/ICSSE.2017.8030871)

[5] Strik, H., Truong, K. P., Wet, F. D., & Cucchiaroni, C. (2007). *Comparing classifiers for pronunciation error detection*. In *Eighth Annual Conference of the International Speech Communication Association*. Recuperado de: <http://citeseerx.ist.psu.edu/viewdoc/downloaddoi=10.1.1.62.614&rep=rep1&type=pdf>

