

Es interesante ver en 2020 los avances, las arquitecturas y las aplicaciones de algo que por años ha estado con nosotros, algo como el reconocimiento del habla (RAH), sin embargo de nada me interesa hablar más que los problemas que ha tenido para su uso y entrenamiento, ya que son muchos.

El reconocimiento del habla consta de una herramienta computacional capaz de procesar la señal de voz emitida por el ser humano y reconocer la información, convirtiéndola en texto, en el desarrollo de esta herramienta son necesarias distintas disciplinas o factores como lo son la fisiología, la acústica, la lingüística el procesamiento de señales, la IA y las ciencias computacionales.

La de a continuación es la fase del desarrollo del RAH, que más me interesa, el aprendizaje ya que son distintas fases de este, pero vemos que en el aprendizaje inductivo se necesita de un factor importante, y es un ejemplo sobre la vida real.

para la decodificación de este ejemplo de vida real, serán necesarias fuentes de información fonéticas, fonológicas y léxicas, a esto se le da el nombre de decodificación acústico-fonética.

luego viene algo muy problemático como lo es el reconocimiento de la gramática, desde mi punto de vista es impresionante el avance en este sentido ya que es sumamente problemático darle a entender a una máquina palabras tan básicas como un “sí” o un “no”, ya que la persona lo puede escribir de una manera u otra, como “sip”, “vale”, “yes”, “okay”, sin contar con las distintas maneras que tiene cada cultura o ciudad de decir “sí”.

esto es solo un ejemplo, pero se pueden generar varios problemas cuando se le entregan palabras a la máquina, ya que si lo que entregamos no tiene sentido para la máquina, por distintos motivos, mucho ruido alrededor, interferencias o acentos cerrados, si la máquina está bien diseñada, pedirá repetir la pregunta al usuario, normalmente cuando se les pide repetir la pregunta, algunos usuarios se frustran o se agitan.

Al día de hoy es común ver herramientas de RAH, pero ¿por qué siguen siendo más utilizados el teclado y el ratón? cuando sería mas rapido hablarle a la máquina y que escriba por nosotros, ya sea buscando algún archivo, o alguna pagina en nuestro navegador, esto se debe muchas veces a que las mayoría de las personas no están bien capacitadas, sin embargo poco a poco es más común utilizar el sintetizador de voz y durante estos años han tenido un gran avance pues existen múltiples herramientas para su uso.

Cada sistema operativo cuenta con una herramienta capaz de reconocer alguna instrucción que le entreguemos en el caso de Windows, tenemos a “cortana”, Google tiene su “Google voice” y finalmente Apple con “Siri”, estos son herramientas usadas para el reconocimiento de la voz, sin embargo ocurren ciertas desventajas usar estas y los distintos sintetizadores, como lo son la imprecisión y lentitud, esfuerzo vocal, factores ambientales, e interpretación de voces inusuales.

Algunas personas no pueden escribir tan rápido como hablan y esto se puede ver como una ventaja para las personas que prefieren rapidez a la hora de escribir textos, esto no siempre es así, ya que cuando dictamos palabras al programa se tienen que hacer correcciones de ortografía, espacios, nombres comunes, una solución algunas veces es agregar ciertas palabras a la biblioteca, pero el proceso de hablarle a la máquina para que escriba por nosotros, se sigue entorpeciendo.

En el 2000 Linda L. Grubbs de la revista PC Worlds informó que este hábito podría ocasionar una lesión en las cuerdas vocales, y aunque no hay vínculo científico entre los programas de reconocimiento de voz y el daño de la voz, se puede pensar en qué pasaría si lo que escribimos en todo un día, lo habláramos pues algún efecto en nuestra voz si podría tener ya sea muy dañino o mínimo, como lo dijo Grubbs eran posible presentar ronquera o tensión.

Para un correcto reconocimiento de la voz también se requiere un ambiente ideal para este, ya que la señal de voz se recibe por medio de un micrófono, este puede captar todo ruido, sonido, palabra, grito, etc.. que no necesariamente venga del usuario de la herramienta es por eso que hablar de los factores ambientales es importantes para un uso correcto, ya que no estar en un ambiente tranquilo o no

tener un micrófono sin filtro, puede presentar un problema para el programa de reconocimiento de voz y terminará escribiendo lo que entienda o nada, es por eso que muchas veces se sigue utilizando el teclado, ya que este no tiene limitaciones en cuanto al ambiente.

Volviendo al tema del entrenamiento, cuando se entrena se la enseñan palabras al programa de reconocimiento de voz, este proceso implica leer varias palabras o pasajes en voz alta, el programa compara tu voz con dialectos regionales conocidos y los ajusta al acento particular automáticamente, algunas personas no alcanzan las figuras de precisión solicitadas por el programa ya que algunas personas tiene acentos inusuales, o acentos muy marcados.

Como dije al principio de este ensayo, son increíbles los avances que han tenido las herramientas del reconocimiento de voz, pero aún quedan cosas por mejorar ya que se presentan muchos problemas para llegar a un uso masivo de estas, cada cultura, cada ciudad de cada país, tiene una manera de decir alguna palabra, tenemos acentos, distintos en especial para la lengua hispana, se pueden presentar más problemas a la hora de la identificación del acento, con el enorme avance de la tecnología, y el desarrollo de esta, veo muy posible que en un tiempo, el uso masivo de las herramientas de reconocimiento de voz, sean de uso totalmente masivo y resuelvan los pocos problemas que hoy se presentan.