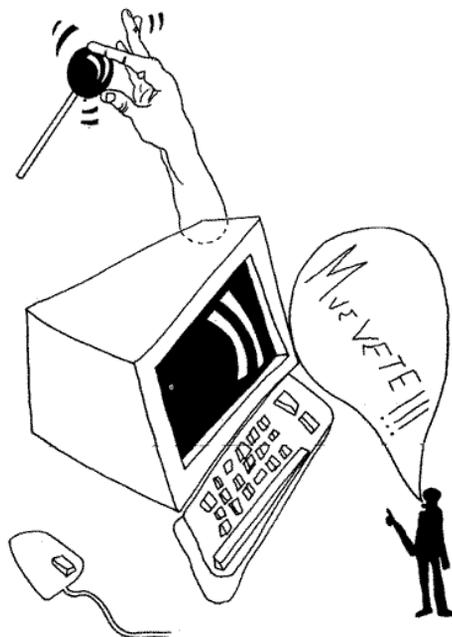


PROCESAMIENTO DE VOZ

Hacia una óptima comunicación hombre – máquina



Uno de los grandes objetivos de la Informática/Electrónica es la capacidad de entablar comunicación hombre-máquina por vía auditiva al estilo de las películas de ciencia ficción. Durante los últimos treinta años, este ha sido el objetivo de miles de investigadores a nivel mundial, siendo el inglés el idioma foco de atención para desarrollar esta tecnología. Muchos han sido los adelantos, en los años 80 la síntesis de voz se logró satisfactoriamente con la comercialización de los conversores *texto-a-voz* capaces de convertir la secuencia de caracteres de un texto escrito en la forma de onda acústica del habla, dotado con una inteligibilidad muy alta y aceptable. El problema inverso, de reconocimiento de voz, es mucho más difícil por la gran variedad de voces, que se traduce en una gran variación en las características acústicas. Parámetros tales como la energía, la velocidad, la entonación, el acento y las condiciones de grabación sufren grandes

cambios entre personas y hasta con la misma persona, dificultando la tarea del computador para tratar de distinguir sonidos correctamente. Sin embargo, en los últimos 10 años se han popularizado dos técnicas sobre todas, una basada en Modelos Marcovianos Ocultos y otra en Redes Neurales, que han permitido obtener un funcionamiento aceptable bajo ciertas restricciones. Varios grados de libertad que dificultan la tarea son el tamaño del vocabulario, la velocidad del habla, el número de usuarios y las condiciones de grabación/cantidad de ruido. Aun los sistemas actualmente comercializados, aunque más desarrollados, tienen limitaciones en su capacidad.

Partiendo de esta problemática el Grupo Procesamiento de Voz, conformado por un equipo multidisciplinario de ingenieros y lingüistas, dirige sus esfuerzos a desarrollar un sistema de reconocimiento automático de ordenes verbales capaz de trabajar por vía telefónica, utilizando como paradigma un sistema de reconocimiento de habla continua. De esta forma, en la etapa inicial del proyecto, se limitan a una cantidad pequeña de palabras, pero en base a una estructura que permitirá su ampliación a un vocabulario extenso de una forma eficiente. A la vez se buscará la adaptación de esta tecnología a condiciones locales de dos maneras: se trabajará con el español y sobre todo con el dialecto venezolano, y se estará trabajando con la red telefónica. La meta es un reconocimiento de 95% de las palabras con habla continuo independiente del usuario, con condiciones de la red y con vocabulario limitado.

Algunas de las hipótesis que maneja este Grupo es que el reconocimiento de voz debe adaptarse a los dialectos locales y a la redes telefónicas locales para mejorar su

funcionamiento, y, que en lugar de entrenar el sistema con voces limpias de ruido, utilizar voces grabadas vía telefónica, a pesar de que este medio tiene una calidad mediocre, buscando aparear las condiciones de entrenamiento con las condiciones de uso a fin de obtener mejor reconocimiento.

Para esto es preciso crear un corpus extenso de voces grabadas por vía telefónica de donde se sacarán las voces para entrenar y las voces para probar el sistema, este debe tener muchos ejemplares de cada palabra que se quiere conocer y debe haber variedad en cuanto a sexo, edad, acento y calidad de grabación; siendo también necesario segmentar las voces grabadas en fonemas; entrenar el sistema al crear modelos acústicos para cada fonema; implementar reconocimiento fonema por fonema; crear vocabulario de palabras a reconocer y establecer su sintaxis definida así como las posibles transcripciones fonéticas de las mismas; probar el sistema; y hacer los ajustes que se consideren convenientes para mejorar el funcionamiento del mismo en función de los resultados que se van obteniendo, entre otros.

Los proyectos actuales del Grupo de Procesamiento de Voz son reconocimiento automático del habla por vía telefónica, reconocimiento de fonemas utilizando Redes Neuronales y Modelos Markovianos Ocultos, colaborar con el postgrado de lingüística en la caracterización acústica del idioma venezolano, colaborar con el Prof. Carlos Gottberg, de la Escuela de Biología, en la caracterización acústica de ranas, y la identificación automática de archivos de audio en sistemas multimedia por vía de su sonido.

Tanto los proyectos desarrollados en el área de ingeniería como los del área de fonética acústica han sido presentados en foros y congresos nacionales e internacionales.