

II. EXPRESIÓN DE EMOCIONES

II.1 Introducción.

La voz es el principal modo de comunicación entre los hombres y consecuentemente se ha estudiado los mecanismos de producción de voz humana y se han creado sistemas capaces de simular y reconocer voz electrónicamente.

Uno de los mayores problemas encontrados en los estudios sobre el habla ha sido el de la variabilidad en ésta. En un gran número de estudios se ha demostrado que varios aspectos del estado físico y emocional del locutor, incluyendo edad, sexo, inteligencia, apariencia y personalidad pueden identificarse solamente por la voz. Todos estos factores, que son diferentes para cada interlocutor, contribuyen a la variabilidad del habla.

La inteligibilidad de los sintetizadores de voz es parecida a la de la voz humana pero estos sistemas no ofrecen esta variabilidad, lo que hace que la voz sintetizada suene “no natural”.

Muchos sintetizadores ofrecen al usuario el control sobre muchas características de la voz humana. Es posible cambiar los parámetros

que controlan la voz para darle diferentes “personalidades” sin afectar seriamente a la inteligibilidad del habla. Si estos cambios en la voz sintetizada reflejan adecuadamente los cambios que experimenta la voz humana cuando expresa emociones sería posible a través de la voz sintetizada simular emociones diferentes.

Para implementar con éxito los efectos emocionales en la voz sintetizada hay que tener en cuenta dos factores fundamentales:

El conocimiento de como pueden distinguirse las características emocionales de la voz y como pueden describirse dichas características usando los métodos de procesado de voz convencionales.

La incorporación de una serie de parámetros o reglas al algoritmo de síntesis de voz, para implementar dichas emociones.

Estudiando la literatura relativa a las emociones en la voz humana y sus efectos puede usarse para diseñar software que controle la calidad del sintetizador de voz, dándole la capacidad de simular emoción en la voz.

II.2. La naturaleza de las emociones.

La emoción no es un fenómeno simple, sino que muchos factores contribuyen a ello. Izard declaró que una definición completa de emoción debe tener en cuenta el sentimiento consciente de la emoción, los procesos que ocurren en el sistema nervioso y en el cerebro y los modelos expresivos observables de emoción.

Las emociones se experimentan a veces cuando algo inesperado sucede y los efectos emocionales empiezan a tener el control en esos momentos.

Para Scherer la emoción puede describirse también como la interfaz del organismo con el mundo exterior, señalando tres funciones principales de las emociones:

- a) Reflejan la evaluación de la importancia de un estímulo en particular en términos de las necesidades del organismo, preferencias, intenciones...
- b) Preparan fisiológica y físicamente al organismo para la acción apropiada.
- c) Comunican el estado del organismo y sus intenciones de comportamiento a otros organismos que le rodean.

Emoción y estado de ánimo son conceptos diferentes: mientras las emociones surgen repentinamente en respuesta a un determinado estímulo y duran unos segundos o minutos, los estados de ánimo son más ambiguos en su naturaleza, perdurando durante horas o días. Las emociones pueden ser consideradas más claramente como algo cambiante y los estados de ánimo son más estables. Aunque el principio de una emoción puede ser fácilmente distinguible de un estado de ánimo, es imposible definir cuando una emoción se convierte en un estado de ánimo; posiblemente por esta razón, el concepto de emoción es usado como un término general que incluye al de estado de ánimo. Más allá de emociones y estados de ánimos está el rasgo a largo plazo de la personalidad, que puede definirse como el tono emocional característico de una persona a lo largo del tiempo.

Muchos de los términos utilizados para describir emociones y sus efectos son necesariamente difusos y no están claramente definidos. Esto es atribuible a la dificultad en expresar en palabras los conceptos abstractos de los sentimientos, que no pueden ser cuantificados. Por ello, para describir características de las emociones se utilizan un

conjunto de palabras emotivas, siendo seleccionadas la mayoría de ellas por elección personal en vez de comunicar un significado estándar.

II.3. Los efectos de las emociones en el habla.

Las primeras investigaciones sobre cómo afectaban las emociones al comportamiento y al lenguaje de los animales fueron descritas brevemente por Darwin en su libro “La Expresión de las emociones en el hombre y en los animales”, publicado en 1872. Más recientemente, los efectos de las emociones en el habla han sido estudiados por investigadores acústicos que han analizado la señal de voz, por lingüistas que han estudiado los efectos léxicos y prosódicos y por psicólogos. Gracias a estos esfuerzos se ha conseguido identificar muchos de los componentes del habla que se utilizan para expresar emociones, dentro de los cuales se consideran los más importantes:

- ❖ el pitch o frecuencia fundamental.
- ❖ la duración.
- ❖ la calidad de voz.

II.3.1. Pitch.

El pitch es la frecuencia fundamental a la que las cuerdas vocales vibran, también llamada frecuencia fundamental o F0. Se considera que las características de la frecuencia fundamental son una de las principales portadoras de la información sobre las emociones.

- ❖ el valor medio del pitch expresa el nivel de excitación del locutor. Una media elevada de F0 indica un mayor grado de excitación.
- ❖ el rango del pitch es la distancia entre el valor máximo y mínimo de la frecuencia fundamental. Refleja también el grado de exaltación del locutor. Un rango más extenso que el normal refleja una excitación emocional o psicológica.
- ❖ las fluctuaciones en el pitch descritas como la velocidad de la fluctuaciones entre valores altos y bajos y si son abruptas o suaves son producidas psicológicamente. En general, la curva de tono es discontinua para las emociones consideradas como negativas (miedo, enfado) y es suave para las emociones positivas (por ejemplo la alegría).

II.3.2. Duración.

La duración es la componente de la prosodia descrita por la velocidad del habla y la situación de los acentos, y cuyos efectos son el ritmo y la velocidad. El ritmo en el habla deriva de la situación de los acentos y de la combinación de las duraciones de las pausas y de los fonemas.

Las emociones pueden distinguirse por una serie de parámetros que conciernen a la duración, como son:

- ❖ velocidad de locución: generalmente un locutor en estado de excitación acortará la duración de las sílabas, con lo que la velocidad de locución medida en sílabas por segundo o en palabras por minuto se incrementará.
- ❖ número de pausas y su duración: un locutor exaltado tenderá a hablar rápidamente con menos pausas y más cortas, mientras que un locutor deprimido hablará más lentamente, introduciendo pausas más largas.

- ❖ cociente entre el tiempo de locución y el de pausas.

II.3.3 Calidad de voz.

La intensidad, las irregularidades en la voz, el cociente entre energías a baja y alta frecuencia, breathiness y la laringerización son algunas de las características que diferencian la calidad de la voz.

- ❖ Intensidad: Está relacionada con la percepción del volumen y se refleja en la amplitud de la forma de onda
- ❖ Irregularidades vocales: Abarcan un gran rango de características vocales. El jitter vocal refleja las fluctuaciones de un pulso glotal al siguiente (como se observa en el enfado) o la desaparición de voz en algunas emociones como la pena, en la que el habla se convierte en un simple susurro.
- ❖ EL cociente entre energía de alta y baja frecuencia: Gran cantidad de energía en las frecuencias altas se asocia con agitación (enfado), mientras que baja concentración de energía en las frecuencias altas se relaciona con depresión o calma (pena).
- ❖ Breathiness y laringerización: reflejan las características del tracto vocal están más relacionados con la personalización de cada voz. Breathiness describe la generación de ruido respiratorio de forma de la componente fundamental tiende a ser más fuerte, mientras que las frecuencias altas son reemplazadas por ruido aspiratorio. La laringerización se caracteriza por una vibración aperiódica de las cuerdas vocales, con un pulso glotal estrecho y pitch bajo, lo que se traduce en una voz chirriante.

II.4. Clasificación de las emociones.

Muchas teorías sobre emociones usan el concepto de emociones básicas, las cuáles son fundamentales, siendo todas las demás emociones modificaciones o combinaciones de estas emociones básicas. Sin embargo, no hay consenso sobre cuáles constituyen las emociones básicas.

Joel Davitz y Klaus Scherer clasificaron las emociones y sus efectos utilizando los ejes o dimensiones de un espacio semántico:

- ❖ Potencia o fuerza: corresponde a la atención – rechazo, distinguiendo entre emociones iniciadas por el sujeto a aquellas que surgen del ambiente (desde el desprecio al temor o la sorpresa)
- ❖ Valencia, agrado o valoración: según lo placentero o desagradable de la emoción (desde la alegría hasta el enfado).
- ❖ Actividad: presencia o ausencia de energía o tensión.

En varios estudios se ha descubierto que se confunden más entre sí las emociones con un nivel similar de actividad (como por ejemplo la alegría y el enfado) que las que presentan similitud en términos de valencia o de fuerza.

También están relacionados el ritmo y la valencia de forma que los sentimientos “positivos” son expresados con un ritmo más regular que los sentimientos “negativos”.

Esto lleva a la conclusión que la dimensión de la actividad está más correlacionada con las variables auditivas relativamente más simples de la voz, como pueden ser el tono y la intensidad, mientras que la

valencia y la fuerza son probablemente comunicados por modelos más sutiles y complejos.

A continuación plantearemos una de las clasificaciones de las emociones y las características de cada una para el idioma inglés que posteriormente compararemos con los resultados obtenidos en nuestro estudio.

El habla neutra suele caracterizarse por un tono con un rango de variación estrecho y unas transiciones de F0 suaves, además de una velocidad de locución alta.

II.4.1. Emociones primarias.

- ❖ **Enfado:** El enfado ha sido ampliamente estudiado en la literatura sobre emociones. Hay contradicciones entre los efectos recogidos en estos escritos, aunque esto puede ser debido porque el enfado puede ser expresado de varias maneras, como veremos en nuestro trabajo. El enfado se define como “la impresión desagradable y molesta que se produce en el ánimo”. El enfado se caracteriza por un tono medio alto (229 Hz), un amplio rango de tono y una velocidad de locución rápida (190 palabras por minuto), con un 32% de pausas.
- ❖ **Alegría:** Se manifiesta en un incremento en el tono medio y en su rango, así como un incremento en la velocidad de locución y en la intensidad.
- ❖ **Tristeza:** El habla triste exhibe un tono medio más bajo que el normal, un estrecho rango y una velocidad de locución lenta.
- ❖ **Miedo:** Comparando el tono medio con los otras cuatro emociones primarias estudiadas, se observó el tono medio más elevado (254 Hz), el rango mayor, un gran número de cambios en la curva del tono y una velocidad de locución rápida (202 palabras por minuto).

- ❖ **Disgusto/odio:** Se caracteriza por un tono medio bajo, un rango amplio y la velocidad de locución más baja, con grandes pausas.

II.4.2. Emociones secundarias.

- ❖ **Pena:** es una forma extrema de tristeza, generalmente causada por una aflicción. Se caracteriza por un bajo tono medio, el rango de tono más estrecho, la pendiente de la curva de tono más baja, una velocidad de locución baja y un alto porcentaje de pausas.
- ❖ **Ternura:** se expresa con un alto nivel de tono que no fluctúa excesivamente.
- ❖ **Ironía:** caracterizada por una velocidad de locución baja y una acentuación muy marcada.
- ❖ **Sorpresa:** con un tono medio mayor que la voz normal, una velocidad igual a la normal y un rango amplio.

Otras emociones secundarias: como el temor, la queja, el anhelo, el aburrimiento, la satisfacción, la impaciencia, el ensueño, la coquetería han sido también objeto de estudio.

Algunos investigadores han utilizado otra clasificación, dividiendo las emociones en:

- ❖ **Activas:** Se caracterizan por una velocidad de locución lenta, un volumen bajo, un tono bajo y un timbre más resonante.
- ❖ **Pasivas:** Caracterizadas por una velocidad de locución rápida, alto volumen, alto tono y un timbre "encendido".

II.5. Aplicaciones de la síntesis con emociones.

Actualmente el uso de sintetizadores para la comunicación de discapacitados se ve afectado adversamente por su artificialidad, convirtiendo la comunicación en una rutina y por la limitación para expresar emociones eficientemente, con lo cual el usuario no puede adaptar la voz sintetizada a las diferentes situaciones y estados de ánimo.

El proveer de una voz personalizada sería una gran ventaja para la ayuda prestada a estas personas. Otra posible utilidad para gente disminuida sería los “periódicos hablantes” y los sistemas de mensajería electrónica con voces elegidas por el usuario para aumentar la inteligibilidad y reducir la fatiga del oyente. Estas técnicas podrían ser también aplicadas en sistemas para personas no discapacitadas, como por ejemplo, las aplicaciones multimedia que se beneficiarían de una mayor variedad de estilos vocales o cabe pensar en un sistema más avanzado, en el que la voz sea el principal modo de interacción con la máquina y el propio ordenador fuera capaz de reconocer cuál es el estado de ánimo del usuario a través de la voz y responder de forma diferente según éste.