

DEPARTAMENT DE DIDÀCTICA DE LA LLENGUA I LA LITERATURA  
DIVISIÓ CIÈNCIES DE L'EDUCACIÓ  
UNIVERSITAT DE BARCELONA

LAS APLICACIONES MULTIMEDIA EN LA DIDÁCTICA DE LA  
PRONUNCIACIÓN DEL INGLÉS EN LA ENSEÑANZA  
SECUNDARIA

por

LLUÍS FIGUERAS HAVIDICH

PROGRAMA DE DOCTORADO: ENSENYAMENT DE LENGÜES I  
LITERATURA (Bienni 1998-2000)

Para optar al título de Doctor en  
Filosofía y Ciencias de la Educación

Director de tesis: Dr. Julià Font i Adrover  
Barcelona 2008

## PARTE II

### 2 MARCO TEÓRICO DEL MULTIMEDIA: La enseñanza asistida por ordenador y su relación con el análisis del habla

#### 2.1 Introducción al proceso audiovisual de comunicación

Al tratar con métodos multimedia es interesante y necesario revisar brevemente el proceso de comunicación audiovisual. Mallas (1977:25) - ver figura 8- comenta la necesidad que los educadores han de tener para conocer el circuito de comunicación cerrado, bajo cualquier fórmula operativa y funcional. Dicho circuito combina los elementos siguientes: emisor, mensaje, receptor, retroalimentación, contramensaje, receptor.

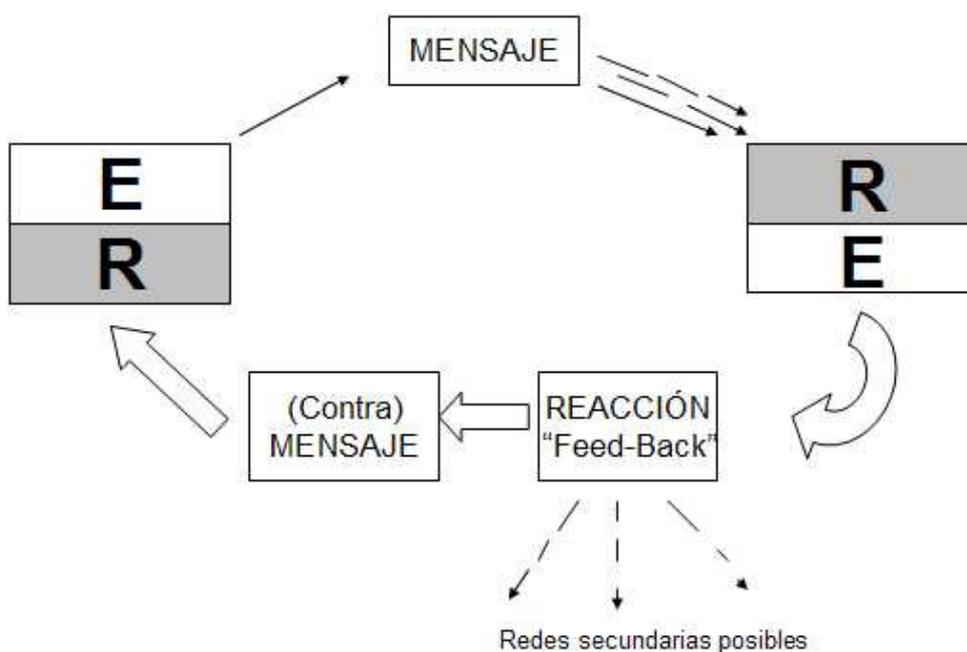


Figura 8. Mallas (1977:25). Circuito cerrado de comunicación

Todo circuito cerrado es bidireccional, pero también omni direccional, ya que en muchos casos se produce una formación de redes secundarias que generan una trama de interacciones que generalmente escapan a todo control. De este tipo de circuitos, Mallas resalta la “retroalimentación” o *feedback* que provoca la inversión de roles comunicativos; es decir, el cambio de flujo informativo y el receptor convertido en emisor. Finalmente, destaca que los audiovisuales que mejor se adaptan a la mayoría de funciones son los *subordinados al profesor* y conocidos también como “pequeños AV” (o medios audiovisuales). El ordenador entraría dentro de esta categoría.

También conviene comentar en qué se basa la irrupción tecnológica y sus necesidades de respuesta teórica, ya que un modelo de enseñanza basado en el uso del ordenador serviría para: enseñar mejor, más rápidamente, a mayor número de sujetos, con mejores instrumentos, paliando la escasez de docentes cualificados y facilitando mejor la tarea docente.

Es importante comentar y observar el llamado Cono de la Experiencia de Dale (1969), que se resumiría en la siguiente cita:

*“El mejor sistema de aprendizaje es, en principio, el de la experiencia. Se aprende y asimila mejor aquello que se hace que aquello que se ve; menos aquello que sólo se escucha, y prácticamente nada, aquello que sólo se oye”.*

Mallas (1977: 29)

En las siguientes páginas podemos observar el gráfico original de Dale (1969) así como un par de interpretaciones llevadas a cabo en varias universidades, que corresponden a las figuras 9, 10 y 11.

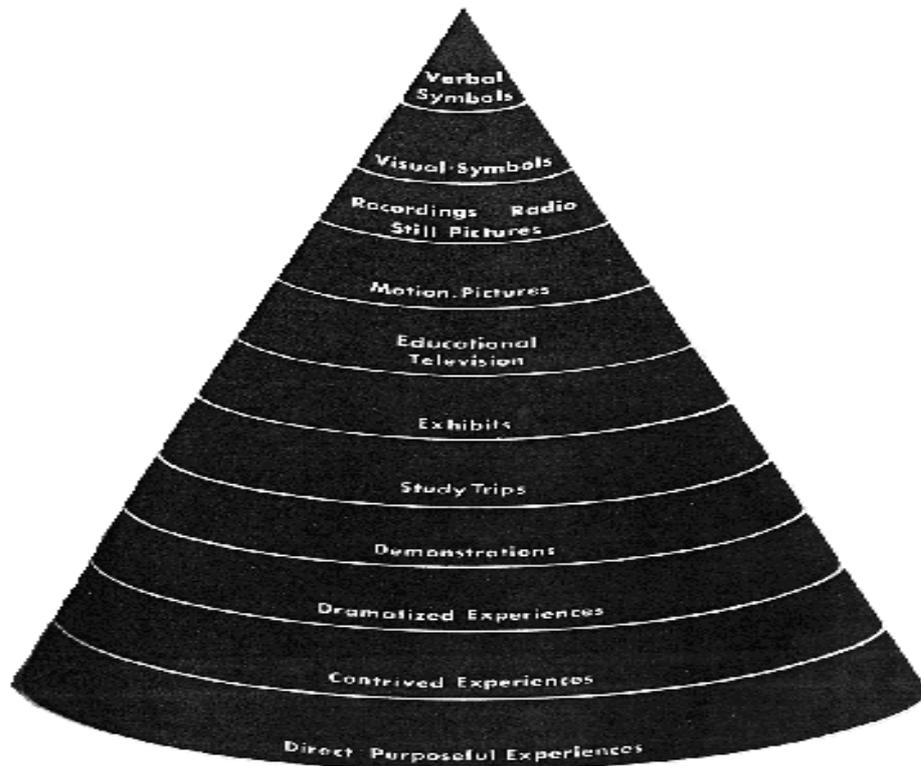


Figura 9. Dale (1969:107). Cono de la Experiencia  
<http://www.work-learning.com/chigraph.htm>

Table VI		
Dale's Cone of Experience		
<b>People generally remember:</b>	?	<b>Levels of Abstraction:</b>
<b>10% of what they read</b>	Read	<b>Verbal Receiving</b>
<b>20% of what they hear</b>	Hear words	
<b>30% of what they see</b>	Watch still picture	
	Watch moving picture	
<b>50% of what they hear and see</b>	Watch exhibit	<b>Visual Receiving</b>
	Watch demonstration	
<b>70% of what they say or write</b>	Do a site visit	<b>Hearing, Saying</b>
	Do a dramatic presentation	
<b>90% of what they say as they do a thing</b>	Simulate a real experience	<b>Seeing and Doing</b>
	Do the real thing	
	? ? ? ? ?	
<small>See Wiman &amp; Mierhenry, Educational Media, Charles Merrill, 1960, for reference to Edgar Dale's Cone of Experience.                      *Question marks refer to the unknown.</small>		

Figura 10. Chacón (2001). Interpretación del cono de Dale  
[http://www.willatworklearning.com/2006/05/people\\_remember.html](http://www.willatworklearning.com/2006/05/people_remember.html)

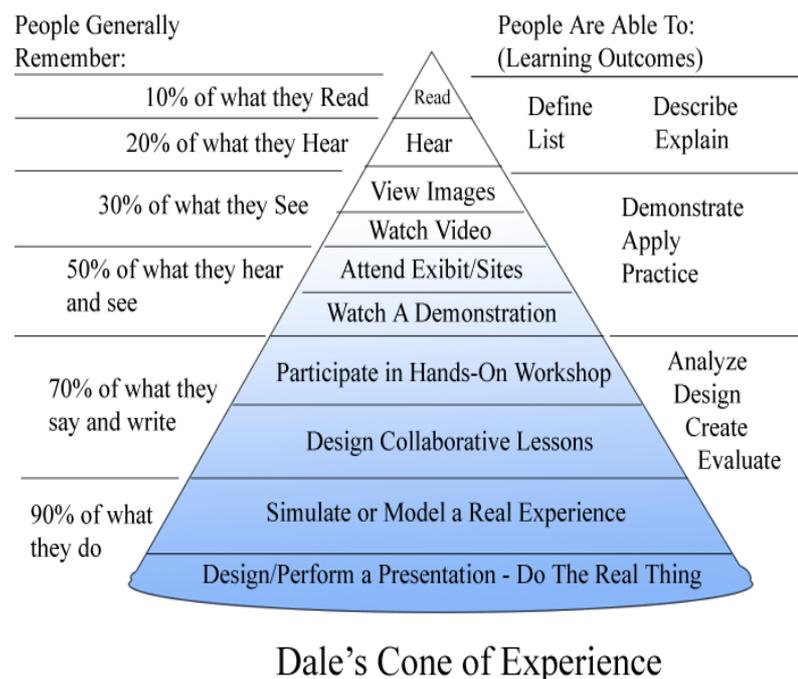


Figura 11. Pastore (2003). Interpretación del Cono de Dale  
<http://teacherworld.com/potdale.html>

La referencia de Chacón (2001), refiriéndose a Wiman y Meierhenry (1969), -ver figura 10- fusionando gráficos y situando números, plantea unos interrogantes que -como el pie de gráfico comenta (en inglés en la página anterior) se refieren al porcentaje inferior al 10% y superior al 90%, cuyo resultado es un punto desconocido.

Mallas (1977) ofrece, por otra parte, una interpretación ligeramente diferente, pero siempre en la línea de los dos gráficos anteriores. Según él, la relación directa con la realidad es mayor a medida que la información del triángulo se dirige hacia el extremo inferior (del 10% al 90%); de la misma manera aumenta la accesibilidad de la interpretación y, por lo que respecta al proceso contrario, aumenta el grado de abstracción y la versatilidad.

## 2.2 Definiciones del término multimedia

El concepto que aquí nos ocupa está relacionado con la irrupción tecnológica de una serie de medios de trabajo y, si bien tiene unos orígenes relativamente recientes, tiene -tal y como refiere Mallas (1977)- unos pre-orígenes *visuales didácticos* encaminados a apoyar la acción magistral (grabados, láminas, encerado, maquetas) que no son aún *medios tecnológicos*, ya que les falta el elemento transmisor: *la "maquina"*, sea un proyector de vistas fijas, etc.

La definición del término multimedia ha evolucionado mucho desde sus orígenes, como podremos observar en diversas citas a continuación. En los años 60 del siglo anterior, Ely (1963) se refería al término multimedia asociándolo con situaciones en las que se combinaban elementos de audio y vídeo de manera múltiple (incluyendo maquinaria considerada hoy en día obsoleta, tales como proyectores de diapositivas, reproductores de vídeo, monitores, cintas de casete, etc.) y otro tipo de *hardware* de una manera instructiva o no instructiva.

Evidentemente, los más de cuarenta años que han transcurrido desde los orígenes de las aplicaciones multimedia, han hecho que el concepto haya expandido su radio de acción y todo ello puede proporcionar al aprendiz una flexibilidad y un uso de estrategias cognitivas.

Alpiste equipara el término multimedia con el punto de encuentro de tecnologías y de "técnicas" que han venido madurando por su cuenta durante años para coincidir ahora en un lugar común y sitúa el efecto pre-multimedia y una referencia económica al coste, en su momento, diciendo:

*“A mediados de los 70 era un privilegio exclusivo de algunas instituciones disponer de algo que hoy nos parece tan elemental como un vídeo y un ordenador. El coste medio de esos equipos, si los comparamos con los aparatos domésticos de hoy, rondaba los dos millones de pesetas<sup>1</sup> (de 1975)”.*

Alpiste (1993:17)

De Bustos (1996), en su guía práctica para usuarios multimedia, se da cuenta de la dificultad de definir el término multimedia de una manera clara, señalando que no existe una definición aceptada por todos. Por tanto intenta hablar del concepto como idea y no como término a definir:

*“Multimedia es básicamente una forma de expresar información, o mejor dicho, son muchas formas de expresar información a la vez, ya que el ser humano es capaz de recibir información de muchas maneras a la vez, tantas como sentidos tiene (...) pero al mismo tiempo es capaz de dar información a cada sentido de muchas formas diferentes, por ejemplo, la vista puede recoger imágenes, colores, movimientos, etc., el sonido puede recibir sonidos armónicos, música, ruidos, etc. La forma de conjuntar todas esas formas de expresión, o las máximas posibles, e integrarlas en un todo, es lo que se conoce como multimedia”.*

de Bustos (1996:15)

Wise (2000) comenta la manera en la que las tecnologías multimedia han evolucionado desde las necesidades militares de producción de herramientas más precisas, y también comenta la problemática al definir el término:

*“Feldman [1997] has defined this sense of multimedia concisely as ‘the seamless integration of data, text, images and sound with a single digital information environment’. Such a definition concentrates on technologies and applications like electronic mail, on-line publishing, CD-ROM, the Internet, and digital television. However, we define multimedia in a less strictly technical way than this. Indeed, as far as the purposes of this book are concerned, such terms as ‘cyberspace’, ‘new media technologies’, ‘computer-based media’, ‘digital technologies’ and ‘information and communications technologies’ are adequate synonyms”.*

Wise (2000:2)

---

<sup>1</sup> Al cambio, 12.000 euros

El paralelismo del término multimedia con la tecnología de la información y la comunicación (TIC), o TAC (Tecnologías del Aprendizaje y del Conocimiento), un término de más reciente creación, así como de otro sinónimo, ‘Tecnología educativa’, serán comentados más adelante. Una definición más precisa, actualizada y desmenuzada en diversos apartados, la encontramos a continuación:

*“Multimedia is the combination of multiple forms of media in the communication of information. Multimedia enables people to communicate using integrated media: audio, vídeo, text, graphic, fax and telephony. The benefit is more powerful communication. The combination of several media often provides richer, more effective communication of information or ideas than a single media such as traditional text-based communication can accomplish. Multimedia communication formats vary, but they usually include voice communication (vocoding, speech recognition, speaker verification and text-to-speech), audio processing (music synthesis, CD-ROMs), data communications, image processing and telecommunications (...) and networks. Multimedia technology will ultimately take the disparate technologies of the computer, the phone, the fax machine, the CD player, and the vídeo camera and combine them into one powerful communication center”.*

Newton (2004: 544)

El gran potencial actual de la tecnología multimedia es la tecnología digital que, tal como podemos ver en el siguiente gráfico de Goodman (2004), figura 12, resume muchos de los aspectos ya tratados:

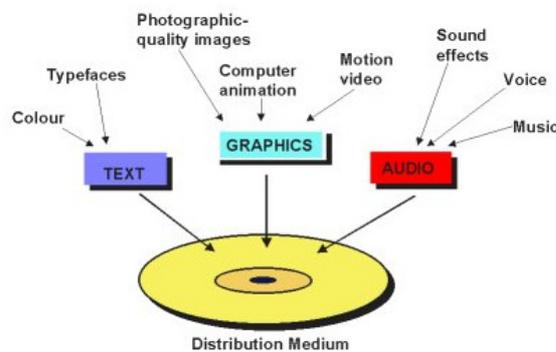


Figura 12. Goodman (2004) Tecnología Multimedia  
<http://www.deakin.edu.au/~agoodman/sci204/lecture1.html>

Alpiste (1993:25) sitúa el primer ordenador personal merecedor del calificativo multimedia al Macintosh II, en su día con sonido y gráficos en 256 colores y con un interfaz amigable. Si bien otros autores, como Wise (2000:44-45), precisan más modelos de maquinarias importantes (Altair –el primer ordenador personal en venderse- y empresas como Apple e IBM PC).

## **2.3 Términos relacionados con Multimedia**

El término multimedia se ha asociado con diversos vocablos y diversos tecnicismos que conviene precisar. A continuación, desarrollaremos términos que merecen su atención, ya que tienen relación con esta palabra, como son: el tutor inteligente, la inteligencia artificial, la tecnología educativa, la interactividad, hipermedia y multimedia interactiva.

### **2.3.1 El tutor inteligente y la inteligencia artificial**

El concepto de *entorno con un tutor inteligente* para referirse al mismo tema podemos definirlo así:

*“An intelligent tutoring environment consists of multiple components including presentation of content material and student interactivity opportunities. The point is that an epistemological framework, instructional theory, and a set of related pedagogical strategies should be at the heart of technology-based instructional design products”.*

Miller (2003: 2)

El hecho de que el constructivismo esté detrás de este *tutor inteligente* hace que se incluyan diversas estrategias formativas, en las que se combinan diferentes aspectos, tales como una aproximación a la

solución de problemas, una inclusión basada en el aprendizaje de un caso, el uso de contextos auténticos, modelos a seguir y un entrenamiento en dichos aspectos, todos ellos conjuntados con el uso de los medios multimedia.

Pennington (1989) utiliza el término *Artificial Intelligence* (AI) para hablar del *software* que hace que el ordenador se comporte de una manera inteligente, comprenda, aprenda y responda. La información restringida a través de la cual opera el ordenador es también conocida como *microworld*. Dicha autora también sugiere los principios de motivación intrínseca, interactividad y eclecticismo y afirma que la cuestión de la comunicación con un ordenador es una de las más intrigantes en CALL.

El llamado Turing Test (ver Saygin *et al.* 2001), basado en el modelo utilizado por Turing en los años 1950 y su pregunta básica acerca de “¿Pueden pensar las máquinas?”, puso en estudio a una persona en conversación con dos mecanismos, uno de los cuales era conducido por un ordenador y otro por un humano. Si no se podía distinguir cuál era operado por el ordenador, se decía que el ordenador había pasado dicho test.

### **2.3.2 La tecnología educativa**

La siguiente cita nos sitúa en un contexto histórico acerca de este tema:

*“La función histórica de la tecnología educativa es más un proceso que un producto. No importa qué nivel de sofisticación puedan tener los métodos de la instrucción, se debe hacer siempre una distinción precisa entre el proceso de desarrollar una tecnología educativa y el uso de ciertos productos o medios dentro de una tecnología de instrucción particular”.*

Saettler (1990: 4)

En relación con la definición de este término, algunos consideran la tecnología educativa como una rama de la teoría y de la práctica educativa que está afectada “primariamente con el diseño y uso de mensajes que controlan el proceso de aprendizaje”.

Ely (1963) hace una clara distinción entre mecanismos y equipos o entre la *ciencia física* de la tecnología educativa y la *ciencia conductista* de la tecnología educativa, y considera *el proceso* como el saber y poder realizar una destreza, y considera el *producto* como el equipo y los materiales utilizados en el proceso. Otra definición de Tecnología educativa sería:

*“Tecnología educativa es un proceso complejo integrado que implica a la gente, los procesos, las ideas, los mecanismos y las organizaciones, para analizar problemas, concebir, implementar, evaluar y gestionar soluciones a esos problemas que implican todos los aspectos del aprendizaje humano”.*

Saettler (1990:6)

Conductistas como Skinner, en la década de los años 1950, concluyeron que eventos mentales tales como pensamientos, imágenes o conocimiento no tenían lugar en la ciencia de la psicología porque no se podían observar directamente. Las posiciones conductistas en psicología empezaron a declinar a finales de los años 50 del siglo pasado, irónicamente a medida que el impacto de la tecnología educativa se aceleraba. Las nociones de Skinner de refuerzo y su aplicación a las máquinas de enseñanza e instrucción programada habían empezado a influenciar el desarrollo de una tecnología de instrucción basada en una ciencia. La instrucción asistida por ordenador (en inglés CAI – *Computer-assisted instruction*) fue influenciada por esta tendencia. La orientación conductista llevaba a un currículum que era programado paso por paso en unidades pequeñas y que hacía hincapié en productos educativos observables.

Por otra parte, el concepto cognitivo de la tecnología educativa, intenta entender los procesos internos de conducta y poner el énfasis en el conocimiento más que en la respuesta. Se ve al aprendiz no como ser pasivo, sino activo, constructivo y participante. Este modelo empezó a reemplazar al conductista a principios de los años 80 del siglo pasado.

Es interesante citar a Richards *et al.* (1998) quienes se refieren a este tipo de estilos definiéndolos como comportamientos psicológicos que sirven como indicadores relativamente estables de cómo los alumnos perciben, se relacionan y reaccionan ante el entorno de aprendizaje; dichos estilos están relacionados íntimamente con tipos distintos de personalidad.

Willing (1988:101) realizó un estudio a gran escala con alumnos de inglés en Australia e investigó cómo las diferencias en estilos cognitivos afectaban las preferencias de los alumnos en seis áreas distintas:

1. Preferencia por ciertas actividades de clase
2. Preferencia por ciertos tipos de comportamiento del profesor
3. Preferencia por ciertos agrupamientos
4. Preferencia por destacar ciertos aspectos de la lengua
5. Preferencia por modos sensoriales concretos, como el aprendizaje visual, auditivo y táctil
6. Preferencia por formas concretas de aprendizaje personal, fuera del aula

Fue precisamente en estos cuestionarios sobre preferencias de aprendizaje donde la preferencia más citada, con un 62%, fue el gusto por practicar los sonidos y la pronunciación.

### **2.3.3 La interactividad**

Otros autores introducen el concepto de interactividad al hablar de multimedia, en la cual el usuario ejerce el control sobre las decisiones a tomar:

*"The most commonly accepted definition of multimedia supports the concept of computer-driven interactivity and a learner's ability to determine and control selection and sequence of content".*

Burton *et al.* (1995: 346)

Matchell y Elliot introducen términos paralelos y relacionados entre sí para hablar de *multimedia* y de *hypermedia*:

*"Interactive multimedia has characteristics of both hypertext and hypermedia. Interactive multimedia should incorporate the following elements: motion video and/or animation; voice, data, graphics, still images, audio; sound and text. Successful incorporation of one or all of these elements can be determined by the degree to which these elements augment, support, and reinforce the achievement of learning objectives and the level of interactivity. However, inappropriate or illogical integration of one or more of the aforementioned media elements may serve to confuse the learner and reduce the possibility of achieving terminal instructional objectives".*

Matchell y Elliot (1991:11)

Más adelante trataremos la interactividad con más detalle.

### **2.3.4 Hipermedia**

La fusión de este término puede ser definida así:

*“The term hypermedia is the merging of multimedia and hypertext. There exists considerable discussion about how both hypertext and hypermedia promote their design as replicating brain functioning in reference to how people learn, access, and use information and ideas. Research has associated relevant psychological phenomena such as cognitive learning principles with the instructional efficacy of hypermedia”.*

Borsook y Higginbotham-Wheat (1992)

Y por lo que respecta a la definición de hipermedia, en la siguiente cita se nombran tres definiciones diferentes, siendo la tercera la más relacionada con los aspectos educativos aquí tratados:

*“A way of delivering information that provides multiple connected pathways through a body of information. Hypermedia allows the user to jump easily from one topic to related or supplementary material found in various forms, such as text, graphics, audio or vídeo”.*

*“Another definition found is: Non-linear media, of which multimedia can be a form. Just as hypertext is a non-sequential, random-access arrangement of text, hypermedia is a non-sequential, random-access arrangement of multiple media such as vídeo, sound and computer data”.*

*“A third definition: Hypermedia is a type of authoring and playback software through which you can access multiple layers of multimedia information related to a specific topic. The information can be in the form of text, graphics, images, audio or vídeo”.*

Newton (2004: 404)

## **2.4 Elementos de un sistema multimedia**

En este apartado vamos a explicar los diferentes componentes que son necesarios en cualquier sistema multimedia y que conviene conocer en lo que respecta a su funcionamiento.

## 2.4.1 Componentes básicos

Alpiste (1993) integra los siguientes componentes básicos de un sistema multimedia:

- Procesamiento multimedia: un microprocesador con cierta capacidad para gestionar el control de imágenes y sonidos.
- Memorias ópticas: todo tipo de procesamientos digitales, entre los que llega a contabilizar más de 20 distintos.
- Interfaz hombre máquina: dispositivos de interacción con los programas. El interfaz sería un vínculo mecánico o eléctrico conectando dos o más piezas de equipos. Existen numerosos dispositivos para la interacción con los programas. Los más difundidos son los teclados y los distintos sistemas de punteo (ratón, pantalla táctil, mandos para consolas o *joystick*, etc). Los lectores de códigos de barras, reconocedores de caracteres y reconocedores de voz son también interfaces.
- Sistemas de salida: donde se presentan los programas.
- Sistemas de entrada: cualquier sistema de captación de imágenes, sonidos o datos que un sistema multimedia pueda aprovechar.
- *Software* de desarrollo: herramientas de “autoría”, programas de dibujo, animación, creación o gestión de recursos multimedia. En el siguiente subapartado podemos ver una ampliación sobre el origen y desarrollo de estos programas, que permiten una flexibilidad adicional en las áreas de:

- Visualización en el monitor: permiten gráficos más atractivos e imágenes.
- Juicio de respuesta: aceptan más de una respuesta.
- Registro de datos (*recordkeeping*): almacenan los diferentes tipos de actuaciones de los alumnos.

- Ramificación de datos (*Branching or looping*): pueden variar los contenidos para usuarios distintos. Algunos programas permiten un convertidor de señal. La tecnología ha de poder producir esa variación de contenidos para también permitir encontrar una actividad y poder volver al punto inicial.

Podemos distinguir cuatro tipos de aplicaciones de aprendizaje con autonomía:

1. Estructurada por el profesor / contenido fijo: normalmente el típico *drill software*, o batería de ejercicios.
2. Estructurada por el aprendiz / contenido fijo: el aprendiz escoge su propio camino. Los alumnos pueden trabajar ejercicios de pronunciación, ver y oír palabras y frases de la lección, grabar su propia pronunciación, escuchar el modelo las veces que quieran y re-grabar para mejorar.
3. Estructurada por el profesor / contenido variable<sup>2</sup>.
4. Estructurada por el aprendiz / contenido variable (ver pie de página 2): conviene también destacar la enorme inversión de horas en este tipo de aplicaciones<sup>3</sup>.

---

<sup>2</sup> Los contenidos variables vendrían a suponer variaciones de contenidos y mayor flexibilidad de trabajo. Si bien en pronunciación, la mayoría de programas estándar suelen formar parte de los dos primeros modelos comentados anteriormente.

<sup>3</sup> Brown (1991) hace una estimación de que por cada hora de material de aprendizaje multimedia se han de invertir entre 50 y 217 horas de trabajo personal de creación de materiales interactivos, si bien no hace ninguna precisión más sobre este dato.

## 2.4.2 El Software de autoría

Antes de cerrar este apartado, debemos hacer una mención a un tipo de programa educativo conocido como “de autoría” (en inglés, *authoring software*) en el cual el contenido es modificable y adaptable a las necesidades del usuario y regulado por el profesor. Aunque en el aspecto que aquí tratamos, como es la enseñanza de la pronunciación, no hay prácticamente ningún programa que se pueda adaptar (a excepción de *Pro-nunciation*), es necesario comentar sus orígenes.

El llamado *Authoring Language* es un lenguaje diseñado específicamente para producir *software* educativo. Según Wachman (1999), éste consiste en programas que el usuario puede adaptar añadiendo datos<sup>4</sup> que encajen con sus necesidades específicas en función del grado de dificultad, interés, cultura, etc., permitiendo un alto grado de autonomía. Davies *et al.* (2005) engloban los mismos usos, añadiendo recursos como imagen o grabaciones de audio y vídeo, que se incorporan como materiales de aprendizaje por medio de plantillas.

Según Gros (1994), el *software* de autoría, aparece a finales de la década de los ochenta del siglo pasado dirigido a diseñadores instructivos, para que no se tenga que utilizar ningún lenguaje de programación. Para ello se han de automatizar algunas de las decisiones que hasta el momento eran realizadas por el diseñador del programa. El diseño instructivo es una tarea compleja y, por consiguiente, su automatización no es un problema trivial.

Gros (1994) distingue tres tipos de automatización: el primero sería la automatización del diseño *y/o producción del software educativo*, la cual responde a la necesidad de cubrir una importante demanda social: el desarrollo de programas de formación de adultos (a distancia, formación

---

<sup>4</sup> Según él, normalmente en forma de texto, aunque por supuesto, con la mejora de los sistemas informáticos es posible incluir audio y vídeo.

en la empresa, etc.). La mayor parte de los programas de formación utilizan tecnología informática sin necesidad de tener que conocer un lenguaje de programación específico. Se trata de automatizar algunas de las decisiones relativas al diseño, la producción del programa, las estrategias de enseñanza, etc. Algunos programas se centran únicamente en la automatización del diseño sin entrar en la producción del programa. Otros, incluyen no sólo el diseño sino también la producción.

Los otros dos sistemas de automatización son: la automatización de *sistemas de transmisión del conocimiento* (tutores inteligentes) y la automatización de *sistemas de ayuda al aprendizaje* (denominados herramientas cognitivas que ayudan a la adquisición de estrategias de aprendizaje).

Estos primeros programas se fundamentaron en el método de producción basado en las teorías del aprendizaje de Gagné (1987), denominado *sistema de desarrollo instructivo*. Este modelo parte de la idea de que, antes de emprender un diseño instructivo, las metas instructivas y los resultados de aprendizaje deben quedar perfectamente definidos en base a cambios observables. El proceso de diseño se divide en cinco etapas: análisis, diseño, producción, implementación y mantenimiento.

En la fase de análisis, el diseñador ha de estudiar los requisitos para la instrucción: contenido, características de los alumnos, niveles de ejecución, etc. En la fase de diseño, hay que desarrollar las especificaciones instruccionales y el test de enseñanza. Mientras que estas fases están dominadas por conceptos abstractos, las dos fases siguientes, producción e implementación, requieren acciones más prácticas. Es necesario crear los materiales de enseñanza y utilizarlos. La fase de mantenimiento permite la evaluación y puesta al día del producto. Es en este aspecto donde la inclusión de palabras, imágenes o de cualquier aspecto modificable entra en acción.

## **2.5 Principios generales que deben regir una aplicación multimedia**

Bou (1997) enumera una serie de principios que deben de regir cualquier aplicación multimedia. Estos son el principio de múltiple entrada, interactividad, libertad guiada, retroalimentación, vitalidad, necesidad, atención y unicidad. Todos ellos son ampliados a continuación en este apartado.

### **2.5.1 El principio de múltiple entrada**

Bou (1997) habla de este tipo de aplicaciones refiriéndose a diseños con perfil de destino, para ser utilizadas por un determinado tipo de usuario. Así pues, distinguimos tres parámetros distintos para almacenar una información en nuestra memoria:

- a) El factor cognitivo: si su complejidad se puede asumir por nuestras destrezas cognitivas.
  
- b) El factor afectivo: el impacto de esa información en nosotros. Además, teniendo en cuenta las teorías de Krashen<sup>5</sup>, también existe otro principio según el cuál los aprendices tienen un *filtro afectivo*, que es una barrera imaginaria que, de manera involuntaria, les impide aprender cuando están estresados, faltos de motivación o inseguros.
  
- c) El factor de la experiencia previa: cómo reaccionamos ante una información similar y en qué esquemas mentales o ideas vamos a integrarla.

---

<sup>5</sup> Brett (1999), en relación a este aspecto, es partidario de dejar explorar libremente a los alumnos aquellas áreas que puede que no les guste admitir que no han entendido.

Por consiguiente, según el principio de múltiple entrada, para lograr una buena comunicación hay que utilizar todos los canales posibles.

### **2.5.2 El principio de interactividad**

Se trata de un recurso muy importante propio de los sistemas informáticos. En consecuencia, el diseño de interacción<sup>6</sup> en una aplicación multimedia debe regirse por unas reglas genéricas que se tratan a continuación:

- a) El refuerzo del mensaje. Podemos incluir quizás en este aspecto lo que Brett (1998b) llama la repetición múltiple, es decir, la posibilidad de repetir el mensaje cuantas veces se desee.
- b) El no utilizar la interactividad supone un desperdicio de potencialidad y de intervención del usuario.
- c) Participación activa, no repetición de gesto. Alpiste (1993) precisa este término aludiendo al ejemplo práctico de utilización de la tecla INTRO para pasar pantallas y, evidentemente, no lo considera como interacción, sino repetición de gesto.
- d) No limitar el mensaje al esquema usuario – máquina: puede promover que las personas dialoguen y cooperen y no debe fomentar un aislamiento personal.

---

<sup>6</sup> Barney y McCabe (2001:4) citan a Vygostky (1962) en relación a la interacción, quien subraya la necesidad de la interacción social para aprender, y mantener esa interacción es clave en el desarrollo del lenguaje del niño. La interacción requiere negociación del aprendizaje y nosotros determinamos lo que nuestros interlocutores saben y necesitan saber para que la comunicación sea efectiva. La interacción no es sólo saber la respuesta correcta o incorrecta, y la interacción real sólo puede existir entre dos seres humanos; por ello el papel del profesor es crucial.

- e) Posibilidad de obtención de un registro de datos descriptivos de conducta del usuario.

### 2.5.3 El principio de libertad (guiada)

El objetivo del guionista es ocultar la estructura interna del programa educativo y que el usuario piense que navega libremente mientras está inmerso en un esquema de etapas predeterminado.

Brett (1998b) habla del concepto de *autonomía* o *del control por parte del alumno*, si bien el principio al que estos dos términos pertenecerían sería el mismo. Healey (1999b) detalla las condiciones que mejoran la autonomía y el aprendizaje en relación al aprendiz (ver figura 13):

- a) Grado de auto-motivación
- b) Preferencia por un estilo independiente
- c) Conocimiento de cómo se trabaja mejor
- d) Conocimiento de cómo uno necesita aprender

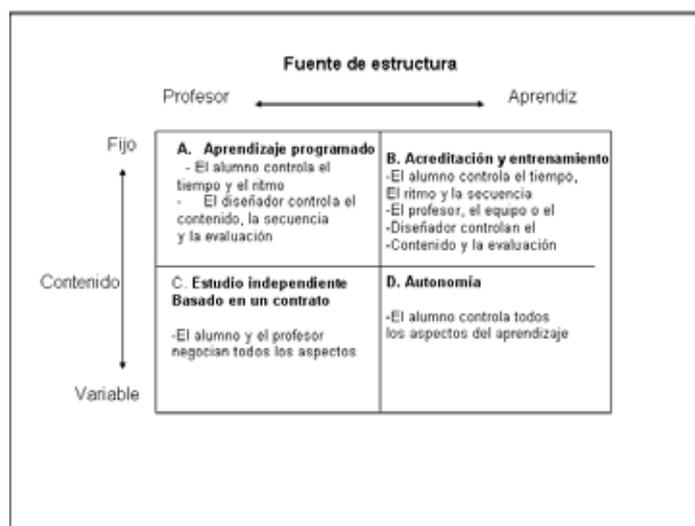


Figura 13. Healey (1999b:393). Parámetros para un aprendizaje autónomo

La figura anterior nos ayuda a ver los diferentes aspectos que se pueden dar dentro de los grados de autonomía entre el profesor y un aprendiz.

#### **2.5.4 Principio de retroalimentación**

Se denomina así a la idea de un sistema que genera información y que se utiliza para corregir su funcionamiento. Para ello hay que tener en cuenta qué tipo de información se recoge, cómo se presenta, a quién se dirige y cómo se procesa. Conviene resaltar más esta idea refiriéndonos a la teoría conductista de adquisición de lenguas de Skinner (1957), según la cual se suponía que el aprendizaje se producía a través de un proceso de retroalimentación o *feedback* de estímulo y respuesta, es decir, imitando el *input* expuesto y reforzándolo. Según esta teoría, la audición tenía un papel relevante en el aprendizaje.

Bou (1997) plantea una idea bastante similar situando lo que él llama la *información elaborada* (ver figura 14 en la página siguiente) como paso siguiente al módulo de análisis de respuestas.

Brett (1999), en relación a la retroalimentación del CALL, dice que este factor está siempre presente, es instantáneo, se aplica a cada respuesta incorrecta, a menudo proporciona la oportunidad al hablante de autocorrección y es privado. En las conclusiones de su tesis doctoral sobre las aplicaciones multimedia y su relación con la comprensión auditiva, Brett concluye diciendo que la retroalimentación de comprensión de una lengua en tiempo real es beneficiosa, así como también el que los aprendices se den cuenta de sus propios errores.

De todas maneras hay que seguir investigando sobre qué tipo de retroalimentación es la más beneficiosa.

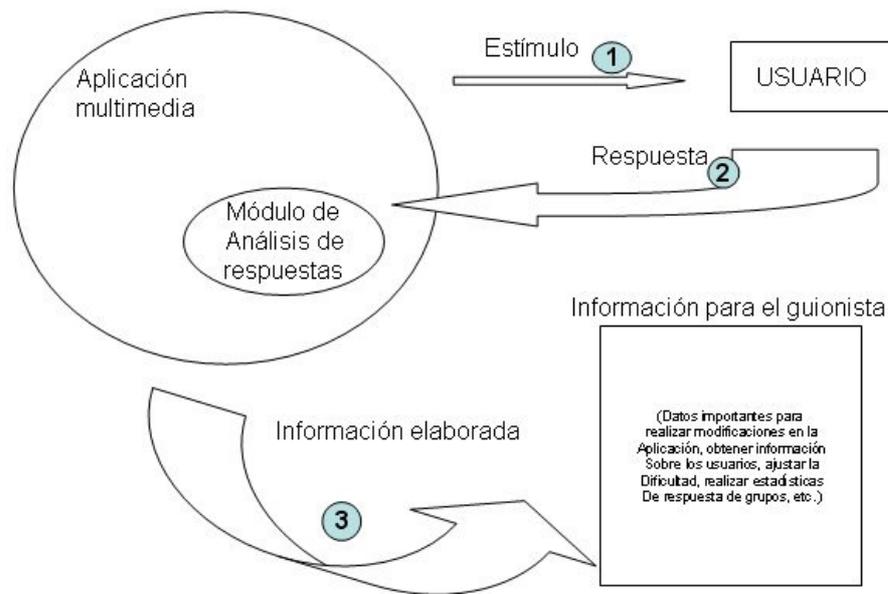


Figura 14. Bou (1997:40). La retralimentación: módulo de análisis de respuestas

### 2.5.5 Principio de vitalidad, necesidad y unicidad

Bou (1997) comenta que toda pantalla está viva. Es decir, el usuario tiene que percibir la aplicación como un mundo al que se asoma. Siempre sucede algo, aunque el usuario no haga nada. Tres consideraciones importantes acerca de este principio servirán para aclararlo un poco más:

- Los iconos animados resultan agradables.
- También lo son los iconos que responden instantáneamente.
- Son desagradables aquellos botones que no van a responder.

A su vez, el principio de necesidad se basa en el hecho de que la aplicación tenga un sentido, que sirva para algo y que sea multimedia. La aplicación viene a resolver un problema cuya solución percibimos inmediatamente que requiere un diseño multimedia. Toda producción

que no nazca de estas dos condiciones es gratuita. Es aquí donde la comodidad y la accesibilidad tienen protagonismo.

Según el principio de unicidad, la misión principal de la persona que coordina una aplicación multimedia es hacer que el producto final parezca obra de una sola persona, que tenga una unidad, una coherencia y un diseño de similares características. Bou (1997) comenta este punto en el sentido de la profesionalidad. Teniendo en cuenta el tipo de aplicación podemos prescindir o no de este principio, pero ha de exigirse una técnica y un estilo.

## **2.5.6 Principio de atención**

El objetivo de las aplicaciones es mantener lo que se llama *la atención sostenida*, es decir, conseguir que el receptor mantenga una actitud continua de expectación ante la aplicación. Para ello se dispone de dos factores descritos en los siguientes puntos.

### **2.5.6.1 La atención cognitiva**

Hemos de tener en cuenta la relevancia de la información y su organización. Ellis (1990:184) resume la perspectiva cognitiva de cómo los alumnos ganan control sobre el conocimiento de otra lengua gracias a la práctica libre que proporciona que el conocimiento de procesamiento controlado se convierta en automático.

Chapelle (1998, 2001) comenta las condiciones que pueden influir en la asignación de la atención durante la realización de tareas en una segunda lengua:

- 1) Interacción modificada: que, según Larsen-Freeman y Long (1991:144), juega un papel importante en la negociación del significado y ayuda a que la entrada de datos sea comprensible mientras aún contiene elementos lingüísticos desconocidos. Han de trasladar al aprendiz a un objetivo y deben centrarse en el lenguaje. La siguiente cita nos resume una buena explicación:

*"These interactions [with the computer] need to move the learner toward a task goal and stop progress along the way to focus on the language. Interaction can be accomplished through mouse clicks and hypertext links or through the learner's linguistic output and the software's speech recognition system. In the latter case, of course, the quality of the interaction will depend in part on the speech recognition capability of the software".*

Chapelle (1998:28)

- 2) Producción modificada: la identificación y corrección de errores dirige la atención a las formas lingüísticas y a la precisión.
- 3) Presión de tiempo: cuando no hay presión de tiempo es probable que exista más atención.
- 4) Modalidad: la comunicación escrita permite más oportunidades para que se forme la atención, mientras que el lenguaje oral suele ocurrir bajo la presión temporal, para que se consiga fluidez.
- 5) Ayuda: cuando los aprendices tienen ayuda sobre algunos aspectos del lenguaje, es probable que tengan más recursos relativos a más aspectos.
- 6) Sorpresa: un elemento sorpresa podría hacer que la atención decreciera, ya que interrumpe los planes y hace que se de importancia a esa sorpresa, pero esta hipótesis dependería de la naturaleza de esa sorpresa.

- 7) Control: el control de varios aspectos de esa tarea por el profesor o el aprendiz puede hacer que se incite a dar más enfoque a la forma, pero se necesita más investigación sobre qué tipo de cuestiones hay que controlar.

### **2.5.6.2 La atención afectiva**

La atención afectiva está basada en el lazo afectivo del usuario con la aplicación. Robinson (1995) define tres tipos de atención:

- Los procesos que están relacionados con escoger la información que se ha de procesar y mantener en la memoria.
- La habilidad de procesar la información.
- El esfuerzo mental en hacerlo.

Schmidt (1990) ve cuatro aspectos de conocimiento relacionado con el aprendizaje de lenguas: intención, atención, conciencia y control. Dentro del campo de conciencia, hace referencia al término “*darse cuenta*” y comenta que se aplica a todos los aspectos del lenguaje, incluida la pronunciación. Aunque esto no quiere decir que automáticamente un ítem del lenguaje vaya a ser recordado y aprendido, sí que este *darse cuenta* es una condición necesaria y suficiente para que el *input*<sup>7</sup> (la entrada de datos) pase a ser *intake* (datos internalizados y adquiridos).

Johnston (1999: 336) se refiere a las responsabilidades del profesor citando a Brown (1994: 93), un autor que ha dado prioridad a temas afectivos, entre los que se incluyen:

---

<sup>7</sup> Este concepto se ampliará en el apartado 2.7.2

- a) Se ha de reconocer la fragilidad del *ego del lenguaje* en una segunda lengua.
- b) La auto-confianza del aprendiz es un elemento crucial en su camino al éxito.
- c) Para triunfar hay que asumir riesgos.
- d) Los encuentros del aprendiz de lenguas con la cultura de la lengua de estudio son una parte vital del proceso de aprendizaje de lenguas.

### **2.5.6.3 La tecnofobia y la ansiedad**

Otro concepto tratado por Johnston (1999) en relación al ambiente sería el concepto de tecnofobia (falta de confianza o desagrado hacia la tecnología, considerada otra forma de ansiedad). Se han tratado también otros conceptos similares, como la *computerphobia*, o ansiedad hacia el ordenador, también llamada *techno-stress*.

Weil y Rosen (1997) se refieren a una serie de estudios y encuestas en las que del 30-80% de la gente muestra algún tipo de ansiedad utilizando ordenadores. Una profundización en el concepto de tecnofobia nos muestra sus características:

- Confusión causada por términos informáticos
- Sentimiento de quedarse atrasado cuando el resto utiliza el ordenador
- Miedo a que trabajar con el ordenador sea demasiado difícil
- Miedo a cometer errores que no se puedan luego corregir
- Una ansiedad más generalizada del papel de los ordenadores en la vida actual
- Problemas diarios en el uso del ordenador: pérdida de datos
- El efecto de deshumanización de los ordenadores
- Sentimiento de falta de control

Existen estudios de Beauvois (1993, 1994) en donde, gracias al uso de la CMC (*Computer-Mediated Communication*) en LAN (*local area network*) en laboratorios y clases de LE, se reducía la ansiedad de la actuación en la clase y se detectó una mayor producción en una clase típica LAN que de la manera oral en clases sin ordenadores. Dichos estudios se llevaron a cabo con un paquete de *software* denominado InterChange.

Young (1999:31) cita a Price (1991) refiriéndose al problema de la pronunciación relacionándolo con la ansiedad en clase. Si bien una pequeña intimidación es necesaria, esta autora en diferentes entrevistas notó que los alumnos parecían muy preocupados acerca de cometer errores en su pronunciación en LE y también deseaban en particular desarrollar un acento que se aproximara al de un hablante nativo. La reticencia a hablar en público era debida principalmente al temor a cometer errores.

### **2.5.7 Estudios con elementos multimedia y aplicaciones con ordenadores**

Según Mayer (1997:14), el aprendizaje multimedia se produce cuando los alumnos reciben información presentada en más de un modo. Al definir el aprendizaje multimedia, es importante distinguir entre *delivery media* (el sistema utilizado para presentar la instrucción), *presentation modes* (formato usado para representar la presente instrucción) y *sensory modalities* (los canales de procesamiento de la información que un aprendiz emplea para procesar la información, tal como el acústico y/o el visual). Mayer predice un efecto positivo en aquéllos que coordinan una explicación visual con una verbal. Los efectos positivos son más importantes en alumnos con bajo conocimiento previo que en aquellos con un alto conocimiento del campo en cuestión. La razón<sup>8</sup> que ofrece es

---

<sup>8</sup> Cabe comentar que en esta investigación, el campo de estudio era el de la mecánica, y se evaluaron actitudes de alumnos que recibieron instrucción multimedia mediante vídeo y audio. El campo que aquí nos ocupa tiene connotaciones diferentes.

que los alumnos de nivel alto es más probable que puedan crear sus propias imágenes mentales cuando se presenta una explicación verbal y que puedan construir conexiones entre las representaciones verbales y visuales.

Brown y Howlett (1994) describen el trabajo de desarrollo y evaluación llevado a cabo en el NCET (National Council for Educational Technology), en Gran Bretaña, que permitió obtener una serie de evidencias. También existían una serie de contradicciones; pero una cuestión quedó clara: no tenía ningún sentido evaluar las contribuciones que puede hacer la TI (tecnología de la información) a menos que los niños y niñas pudieran trabajar con ordenadores de manera regular. A continuación podemos resumir una serie de cuestiones que sirven para ver el valor de la TI en el entorno educativo:

- 1) La TI puede proporcionar un entorno de aprendizaje seguro: el miedo a equivocarse obstaculiza los primeros pasos, pero experimentar con nuevas ideas y aprender de los errores es importante. El ordenador proporciona un entorno seguro. Permite detenerse a pensar sobre un problema sin molestar al resto de la clase. El ordenador no se cansa y no es impaciente ni ridiculiza los esfuerzos del estudiante.
  
- 2) La TI tiene flexibilidad para adaptarse a las capacidades y necesidades individuales de cada estudiante: puede trabajar a la velocidad adecuada adaptada a las necesidades de cada estudiante en vez de seguir un ritmo preestablecido. Puede proporcionar diversos puntos de entrada y salida, no juzga e informa al estudiante de su éxito o fracaso sin decir si el estudiante es bueno o malo.

- 3) La TI puede estimular a los estudiantes desmotivados: la información puede ser más interesante, se trabaja con un ritmo propio. Ver en privado que no se ha entendido algo hace que sea más fácil aceptar ayuda.
- 4) Los ordenadores proporcionan a los estudiantes la oportunidad de conseguir buenos resultados donde previamente han fallado.
- 5) Los ordenadores pueden reducir el riesgo de fracaso escolar.
- 6) La TI puede presentar la información mediante formas nuevas que ayuden a comprenderla, asimilarla y utilizarla más eficazmente.
- 7) La tecnología interactiva motiva y estimula el aprendizaje.
- 8) Los programas informáticos con voz digitalizada pueden ayudar en el proceso de lecto-escritura.
- 9) Los ordenadores ayudan a los estudiantes a aprender cuando se utilizan tareas y actividades significativas y bien diseñadas: tienen resultados claros de aprendizaje, se describen de manera clara y sencilla, ofrecen oportunidades para diversos estilos de aprendizaje y son atractivas. Los estudiantes han de saber de qué manera se espera que hagan su trabajo. Han de saber: qué aprenderán, por qué lo han de aprender, si la actividad es individual o en grupo, cuándo tendrán acceso al ordenador y cómo será evaluado su trabajo.
- 10) El uso de la TI hace que los profesores revisen cómo enseñan y cómo aprenden los estudiantes: los estudiantes pueden investigar por sí solos y el profesorado se dedica a proporcionar nueva información, a interrelacionarla con conocimientos adquiridos o redirigir el pensamiento cuando el aprendizaje se bloquea.

## **2.6 Otros aspectos importantes**

Podríamos señalar diferentes aspectos generales que son reconocidos en el aprendizaje de lenguas y que tampoco conviene olvidar en relación con el tema que nos ocupa.

### **2.6.1 La motivación y el interés**

Brett (1999) y Watts (1989) encuentran evidencia en el uso de ese tipo de aplicaciones para trabajo autodidacta por encima del que pueden ofrecer libros, cintas o vídeos.

Skehan (1989) revisa la investigación de la motivación y también de diversos otros factores, tales como la aptitud, estrategias para el aprendizaje de lenguas y factores cognitivos y afectivos, que más tarde los subdivide en: ansiedad, independencia de campo, toma de riesgos, inteligencia y extroversión / introversión.

Bernaus (1992) comenta su punto de vista del proceso de aprendizaje y adquisición de una lengua extranjera diciendo que el lenguaje constituye un proceso único, donde el factor social puede jugar un factor fundamental.

Setter y Jenkins (2005:1) afirman que es a nivel afectivo a través de la manera por la cuál hablamos y, especialmente mediante nuestro acento, donde proyectamos nuestra identidad étnica, social y regional.

En sus referencias a la motivación, Skehan (1989) formula varias hipótesis sobre este fenómeno:

- La motivación intrínseca produce un interés en la tarea del aprendizaje.
- La motivación resultativa incrementa o reduce la motivación según exista éxito o fracaso en una tarea.
- La causa interna, que es la motivación que los alumnos traen con ellos en la situación de aprendizaje.
- La presión externa influenciará la motivación.

Good y Brophy (1987), al referirse a tecnología y aprendizaje, sugieren cinco condiciones iniciales para motivar que deben llevarse a cabo por el profesor como *facilitador* en la clase:

- a) Un apropiado nivel de desafío o de dificultad
- b) Objetivos de aprendizaje que sean significativos para el alumno
- c) Variación en los métodos de uso
- d) Retroalimentación intrínseca (a nivel muy primario, utilizando el procesador de textos, puede dar al alumno un cierto grado de satisfacción) y extrínseca
- e) Inexistencia de barreras para el aprendizaje

Brett (1999) y Crookes y Schmidt (1992) vieron cinco dimensiones de clase que pueden afectar o ser afectadas por factores que están en relación con la motivación:

- La fase preliminar de la clase
- Las actividades de clase
- El *feedback* o retroalimentación
- Los efectos de la percepción de los estudiantes y de los materiales

El hecho de que los alumnos hayan expresado su interés en ciertas áreas, por ejemplo la pronunciación, puede hacer que ciertas actividades en este campo se puedan integrar. Según las conclusiones de la tesis doctoral de Brett (1999) sobre el multimedia como medio, éste no proporciona una motivación a largo plazo. Sin embargo, cuando la motivación para usar la aplicación era extrínseca, las actitudes no fueron positivas. La aplicación no tuvo la habilidad de motivar al aprendiz hasta el punto de usarla cuando no fuera estrictamente necesaria.

### **2.6.2 Multimedia y niveles de motivación**

Situándonos en el término que Brett (1999) designa como un *micro nivel*, el multimedia puede incrementar la atención y ayudar en el desarrollo de estrategias útiles para el aprendizaje de lenguas:

*“Learners’ motivation to deploy useful learning strategies might also be beneficial to acquisition. Multimedia-delivered materials may well be capable of providing the forum for supporting, encouraging, or motivating the use of metacognitive strategies such as “monitoring” and being able to evaluate one’s own performance. Directed attention may be encouraged through meaning-focused tasks with feedback. It might be that in the multimedia environment such metacognitive strategies can be used more easily, or conversely, it might be that it is the multimedia environment itself which motivates the use of such strategies”.*

Brett (1999:14)

Según él, el hecho de que se pueda configurar el entorno multimedia permite que haya una cierta flexibilidad y se pueda acomodar con estudiantes de diversas interlenguas, diferentes estilos cognitivos, diferentes estilos de aprendizaje y diferentes necesidades. He aquí los contenidos que ofrece el multimedia educativo según Brett:

- Aprendizaje del contenido: lo que se puede ver, lo que se puede escuchar y qué cantidad.
- Modo de aprendizaje: vídeo, audio, con o sin subtítulos.
- Activo o pasivo: completar la tarea o no.
- Tipo de tarea.
- Nivel de dificultad en niveles o escalonado.
- Secuencia de aprendizaje: qué hacer, cuándo y en qué secuencia.
- Tiempo y ritmo de aprendizaje.
- Consejo en aprendizaje de lenguas y manera de explotar el programa a través de Internet.
- Materiales del mundo real: a través de vídeos, CD-Roms, audiciones, etc.

### **2.6.3 Condiciones para entornos de aprendizaje de lenguas**

Diversos autores han propuesto diversos modelos de entornos de aprendizaje, como Warschauer (1996), Chun y Plass (2000).

Egbert *et al.* (1999: 3-6) se refieren a una serie de condiciones óptimas de aprendizaje de lenguas:

1. Los alumnos han de tener oportunidades para interactuar y negociar el significado.
2. Los alumnos están involucrados con tareas auténticas.
3. Los alumnos están expuestos a, promocionan el uso de y producen lenguaje variado y creativo.
4. Los alumnos tienen suficiente tiempo y retroalimentación.
5. Los alumnos están guiados para atender conscientemente el proceso de aprendizaje.
6. Los alumnos trabajan en un ambiente con un nivel ideal de stress y ansiedad. Brett (1998a) ha utilizado unos términos similares,

tales como *libre de juicio y libre de peligro*, que se pueden relacionar con este punto.

7. Los alumnos tienen un nivel de autonomía.

#### **2.6.4 Desventajas en el uso de las aplicaciones multimedia**

Brett (1999) se refiere al coste y al equipamiento como el principal problema. El coste de los equipos va bajando progresivamente, pero las novedades y el hecho de que los equipos se vuelvan obsoletos es un grave problema a tratar.

Edwyn James<sup>9</sup>, del Centro para la Innovación y la Investigación de la Educación de la OCDE (Organización para la Cooperación y el Desarrollo Económico) afirmó en el año 2001 que la mayoría de los ordenadores tenían un período de vida de cinco años, en el mejor de los casos. James se preguntaba cómo se podía justificar una inversión de grandes cantidades de dinero en *hardware* en una escuela que sólo abría seis horas al día, cuarenta semanas al año, si en cinco años el material estaría obsoleto. Las posibles soluciones pasaban por permitir el acceso al resto de la comunidad, lo cual facilitaría y rentabilizaría esas inversiones.

Los recursos informáticos y audiovisuales de una escuela pueden no satisfacer a todos. Si bien cada vez más el número de personas interesadas en trabajar con ellos aumenta, hay que trabajar en la formación de profesores a nivel informático para superar los obstáculos con los que se pueden encontrar, como son la solución de pequeños problemas técnicos y la mejora en la búsqueda de sus propios recursos didácticos y planificación escolar.

---

<sup>9</sup> En Fine (2001)

El cansancio en relación a la dificultad de lectura en un ordenador ha sido tratado por Thibodeau (1997), que afirma que es más cansado, más difícil y un 28% más lento el leer en un ordenador que leer en formato papel.

Milton (1997) indica que el problema con la aplicación de materiales en CD-Rom que dependen de becas y ayudas gubernamentales, europeas o privadas es siempre el mismo, la falta de continuidad o el excesivo coste del proyecto. Proyectos como “Tune in English<sup>10</sup>”, cuyo prototipo fue mostrado en unas jornadas de fonética y que contribuía especialmente a la enseñanza de la prosodia, nunca vio la luz quizás debido a estos problemas.

### **2.6.5 Desventajas lingüísticas en diversos estudios sin medios multimedia**

Diversos autores se han referido a problemas muy concretos observados en actividades lingüísticas que podemos relacionar a continuación. En la introducción de su libro sobre la audición, White (1998) enumera una serie de problemas que se producen en la clase de lengua extranjera:

- a) No existe mucho tiempo para escuchar una cinta.
- b) No se dedica mucho tiempo para analizar lo que se contestó mal.
- c) Frecuentemente, la clase de audición hace que los alumnos sean sujetos pasivos.
- d) Las tareas no tratan de vincular la audición con el habla.
- e) A menudo se espera el 10% de la comprensión.

White (1998) cita a Bone (1988), el cual, en un estudio con hablantes nativos, indicó que la gente a menudo escucha un 25% de su potencial e ignora, olvida, distorsiona y malinterpreta el restante 75%.

---

<sup>10</sup> Kaltenboeck (2001)

La concentración aumenta un 25% si ellos consideran que lo que están oyendo es importante y están interesados, pero nunca alcanza el 100%.

## **2.6.6 Los estudios de la memoria**

Un apunte importante en relación a la memoria humana debe considerarse en esta tesis. Si bien este no es uno de los pilares de trabajo y puede abordarse en posteriores tesis, es necesario hablar del procesamiento de la información humana y poder establecer una cierta relación con los métodos multimedia.

### **2.6.6.1 La Psicología cognitiva y los ordenadores**

Navarro Guzmán (1993:173) considera que, con el desarrollo de la Psicología Cognitiva, los procesos mentales vuelven nuevamente a ser objeto de interés de la Psicología científica, debido fundamentalmente al auge de los ordenadores, que analógicamente con el hombre, reciben, almacenan y recuperan información siguiendo estrategias predefinidas. Por esta razón, se produce una incorporación masiva de términos cibernéticos, tales como: almacenamiento, codificación, búsqueda, recuperación y un desplazamiento del enfoque asociacionista del estímulo-respuesta, dando paso a un enfoque cognoscitivo y experimental. Los términos codificación, retención y recuperación se utilizan para definir los tres aspectos básicos de los sistemas de memoria que tienen lugar bajo el control del sujeto. La codificación se refiere a la manera en que se almacena o representa la información; la retención, a la conservación de esta información durante un período de tiempo más o menos prolongado y, la recuperación, a todos aquellos procesos que hacen posible su utilización posterior.

El mismo autor, al referirse a la neurofisiología de la memoria y el aprendizaje, afirma:

*“El aprendizaje es el proceso de adquirir un conocimiento sobre el mundo, es decir, la adquisición a través de la experiencia. La memoria es la retención o almacenaje de ese conocimiento, así como su evocación o consecuencias”.*

Navarro Guzman (1993:51)

Prácticamente no hay actividad nerviosa que no influya o no sea afectada en alguna medida por el aprendizaje y la memoria, ya que la segunda vez que hacemos o percibimos algo, nunca es igual a la primera.

Saettler (1990: 323), en referencia a los ordenadores, citaba a Neisser (1967) diciendo que dichas máquinas proporcionaron a los psicólogos la garantía de que los procesos cognitivos fueran tan reales como las respuestas musculares y glandulares de la conducta humana.

### **2.6.6.2 El modelo de memoria Atkinson-Shiffrin**

Si consideramos uno de los modelos de memoria más aceptados en la actualidad<sup>11</sup>, el modelo Atkinson-Shiffrin, propuesto en 1968, considera que la memoria tiene tres almacenes: los registros sensoriales (incluyendo memorias icónicas y ecoicas) o el almacén de memoria sensorial a corto plazo (STSS), la MCP (almacén de memoria a corto plazo), a veces llamada también memoria de trabajo (WM) y la MLP (a largo plazo). La información en el STSS decae rápidamente (al cabo de de un segundo) y se perdería totalmente si no se transmitiera a la MCP, que tiene una capacidad pequeña y la información decae en un tiempo de 5 a

---

<sup>11</sup> La referencia que se hace hacia el modelo Atkinson-Shiffrin (1968) en relación a uno de los modelos de memoria ha sido la que se ha encontrado como más empleada al consultar diversos autores (Miller, 1956; Peters y O'Sullivan, 2003; French y Anselme, 1999). Otro modelo, como la Teoría de los Niveles de Procesamiento ( Craik y Lockhart, 1972) no se considera como un modelo rival o en desacuerdo con el que aquí definimos.

18 segundos. Una de las características de la MCP es el mecanismo de ensayo, que se ve parcialmente bajo el control del individuo, pero con una capacidad de procesamiento fijo. Ya que la MCP tiene capacidad limitada<sup>12</sup>, la información no transferida directamente a la MLP se pierde.

Navarro Guzmán (1993:191), en relación a la MLP, comenta que el principal problema de este almacén de memoria es cómo recuperar la información que en él se encuentra. El acto de recordar sería, en este sentido, la aplicación sistemática de reglas para analizar esta información almacenada, utilizando estrategias y mecanismos mentales que tienen mucho que ver con el pensamiento y la asociación de problemas. La recuperación va a depender, en gran medida, del papel crucial que tenga la organización en el momento del aprendizaje.

El proceso de memoria sensorial empieza con señales físicas (visuales, acústicas o táctiles) que se reciben por los registros sensores (ojos, oídos y piel). Neisser (1967) ha llamado al input visual como “icono”, al auditivo como “eco” y al tercero como “táctil”.

En relación al modelo Atkinson-Shiffrin y el progreso en la teoría de la memoria, la siguiente cita se refiere a las diversas posibilidades de permanencia de ítems en la memoria:

*“...Un aspecto fundamental en el modelo Atkinson y Shiffrin es que cuanto mayor es la permanencia de un ítem en la MCP, mayor es la probabilidad de que dicho ítem pase a la MLP. Como el único modo de aumentar la permanencia de un ítem en la MCP es repasarlo, cuantas más veces se repase, mayor será la probabilidad de que se transfiera a la MLP. Ahora bien, el paso de un ítem a la MLP no garantiza plenamente su recuperación, puesto que cuando deja de repasarse en la MCP empieza a debilitarse su huella en la MLP”.*

Navarro Guzmán (1993:233)

---

<sup>12</sup>La evidencia sugiere que la MCP es muy limitada y que se ha de activar en ella sólo un pequeño número de ítems de información en un tiempo establecido. En Saettler (1990:325), se sitúan esos ítems en unos siete objetos, o siete unidades de información, y que para retenerlos, la única opción es el ensayo.

### 2.6.6.3 Estudios sobre los efectos de la memoria

En los años 1960, en los laboratorios de teléfono Bell -encabezados por George Sperling- se realizaron diversos estudios visuales y auditivos, utilizando una técnica conocida como *el informe total*, según la cual se puede controlar cuántos estímulos (letras, palabras, dígitos) son captados de un solo ‘golpe de vista’.

Unos años más tarde, Darwin *et al.* (1972) también investigaron la memoria auditiva y vieron que la duración temporal del almacenamiento auditivo era generalmente más larga que la del almacenamiento visual.

Penney (1989) recuerda la división del almacenamiento de la información presentada según sea por vía oral o visual al hablar de la retención de material verbal en la memoria a corto plazo, y según su estudio previo de 1975, dice:

*“If there are separate processing streams for auditorily and visually presented items in short-term memory, it should be possible to show that retention is greater if information is presented to two modalities than when only one mode is used”.*

Penney (1989:401)

Gathercole *et al.* (1992), en sus estudios sobre la memoria de trabajo, recogen brevemente el estudio de Service (1992), que evaluó las habilidades de repeticiones de palabras que no tenían significado con un grupo de niños finlandeses de nueve años que aprendían inglés como lengua extranjera, considerando los progresos como muy positivos y fueron relacionados dos años y medio después. Este autor argumenta que la asociación entre la repetición de palabras sin significado y los logros de los niños en el aprendizaje de lengua extranjera están mediatizados por la contribución de la memoria de trabajo fonológico en la adquisición de vocabulario.

En relación al tratamiento de la pronunciación, debemos reflexionar acerca de los aspectos descritos anteriormente, ya que no debe tratarse de un recurso aislado y para fines única y exclusivamente orales. El hecho de poder trabajar con elementos multimedia permite integrar magníficamente tanto los aspectos orales como los visuales y establecer relaciones sobre vocales y consonantes que equivalen o se asemejan a sonidos (largos y cortos), dado que la integración del Currículo educativo contempla diversas áreas de interés (comprensión, expresión, orales y escritas).

Brett (1999), en su capítulo sobre el desarrollo del control en el uso del conocimiento de la L2, habla del modelo de procesamiento de la información de McLaughlin (1987). Este modelo tuvo un impacto en la pedagogía del aprendizaje de lenguas. Este autor observa dos maneras de procesar la información en relación con la entrada de datos: el procesamiento controlado y el procesamiento automático. Los dos modos de procesamiento requieren la participación del conocimiento que ya ha sido adquirido y almacenado en la memoria como nodos. Para facilitar la comprensión acerca de lo que es un nodo, Brett (1998b) cita a Shiffrin y Schneider (1977) para referirse a la memoria, diciendo que...:

*“... es una gran colección de nodos que pasan a estar ínter asociados de manera compleja a través del aprendizaje. Cada nodo es un grupo o un conjunto de elementos de información. Muchos de los nodos son interactivos o pasivos y, cuando están en este estado, el sistema interconectado de nodos se llama Memoria a largo plazo”.*

Brett (1998b: capítulo 2.4.4. online)

El segundo modo de procesar la información es el de manera automática; éste ocurre cuando la entrada de datos (*input*) específica de L2 se activa y se conecta con un conjunto de nodos. Este conjunto conectado de nodos se ha formado como resultado del aprendizaje.

El procesamiento automático es el resultado de la práctica utilizando el modo de procesamiento controlado. Gracias a la participación en actividades en tiempo real, los alumnos tienen la oportunidad de utilizar estos procesos cognitivos que automatizan el conocimiento.

#### **2.6.6.4 El papel de la música y la canción en relación a la memoria**

Murphey (1992: 6-7) habla de la importancia de la música y de la canción en el aprendizaje de lenguas. A continuación se citan algunos aspectos que consideramos importantes:

- a) El fenómeno de “la canción *que se repite incesantemente en mi cabeza*” parece reforzar la idea de que las canciones trabajan nuestra memoria a corto y a largo plazo<sup>13</sup>.
- b) La recitación de canciones se parece a lo que Piaget (1923) describió como lenguaje egocéntrico, en el que los niños hablan sin prestar atención a nada en especial. Simplemente disfrutan con repetir. Podría decirse que la necesidad de ese lenguaje egocéntrico nunca nos abandona y se lleva a cabo parcialmente a través de la canción. Krashen (1983) ha sugerido que esta repetición involuntaria puede ser una manifestación de lo que Chomsky denominó “el mecanismo de adquisición de lenguas”. Parece que nuestros cerebros son propensos de una manera natural a repetir lo que oyen en el ambiente para poder darle sentido. Las canciones pueden activar de manera importante el mecanismo de repetición de la adquisición de lenguas, donde también podemos integrar la pronunciación. Ciertamente parece que es así con los niños, que aprenden casi sin esfuerzo.

---

<sup>13</sup>Conceptos citados en Murphey (1992) tomados de su libro anterior (1990b) *The song stuck in my head phenomenon: a melodic din in the LAD? System*18/1, 53-64. Consideramos importante recalcar dichos aspectos, ya que existen prácticas derivadas para trabajar pronunciación en producciones musicales.

#### **2.6.6.5 El ensayo mental en pronunciación. Hallazgos**

Barrera-Pardo (2004), en su estudio relacionado con la precisión, recogió información relevante en tres áreas del aprendizaje de la pronunciación que normalmente escapan a la observación experimental directa. Barrera-Pardo comenta el uso de tres estrategias diferentes:

- 1) La evaluación mental: estrategia usada por el aprendiz que no implica ensayo alguno de lo ya estudiado, sino que simplemente hace reflexionar al alumno sobre lo trabajado.
- 2) El ensayo mental espontáneo: ya tratado en Krashen (1983), que describió el ensayo involuntario de las estructuras de L2 (palabras, sonidos y frases) que hemos tratado en el punto anterior.
- 3) El ensayo encubierto en pronunciación: un tipo de práctica voluntaria en la que el aprendiz practica (repitiendo, recordando o imitando) lo que ha aprendido. Esta estrategia sirve para diferentes propósitos, tales como una herramienta de autocorrección, una manera de sintonizar con material ensayado y también como rastreo del habla al que el aprendiz está expuesto.

Dickerson (1994) conceptualiza el ensayo encubierto de la pronunciación como una estrategia de aprendizaje activa, cognitiva y de auto-evaluación, que requiere cierto esfuerzo e implica un uso consciente de reglas y otros recuerdos de producciones orales, que pueden repararse y practicarse hasta mejorar y sentirse bien. En su experimento con dos grupos (uno con alumnos de pronunciación avanzada y otro con menor nivel en pronunciación), las conclusiones del estudio indicaron que los alumnos de mayor nivel mostraban un uso superior de estrategias, y que

en general practicaban más de manera encubierta sus propias autocorrecciones y accedían a ellas- que los alumnos de menor nivel.

Dichos hallazgos sugieren que la evaluación mental del habla y la atención a la práctica encubierta de la pronunciación pueden ser la característica de los hablantes más avanzados.

## **2.7 El aprendizaje de lenguas asistido con ordenadores (CALL)**

Los diferentes libros que hacían referencia al CALL empezaron a publicarse en la década de los años 1980. Si bien éste es un campo aún inmaduro<sup>14</sup>, el uso de ordenadores en la instrucción del lenguaje está basado en la fusión de grandes tendencias en la educación de los años 50 y 60 del siglo pasado. Es necesario referirse al Audiolingüismo, ya que este método fue una aproximación teórica y metodológica al aprendizaje de lenguas -también con raíces conductistas- que estaba basado en la formación de hábitos y se implementó para ser utilizado en un programa de aprendizaje con un ordenador o también en un laboratorio de idiomas.

Según Larsen-Freeman (1986), el método audio lingual se desarrolló en los EEUU durante la Segunda Guerra Mundial, en un momento en que era necesario aprender una lengua extranjera debido a propósitos militares. Mientras el objetivo de comunicación en la lengua de estudio era el objetivo del método directo, existían en ese momento nuevas ideas acerca del lenguaje y del aprendizaje que emanaban de las disciplinas de

---

<sup>14</sup> Hubbard (1996) dice que el CALL no es un campo maduro, si lo fuera, podríamos definir la metodología CALL como un conjunto de métodos empleados por los profesionales de CALL, donde el método representa un conjunto organizado de principios consistentes y su realización es una parte de un curso educativo o una actividad basada en ordenadores.

la lingüística descriptiva y la psicología conductista. Dichas ideas contribuyeron a desarrollar el método audio lingual.

A nivel tecnológico, la invención de los laboratorios de idiomas prometía la realización de principios de aprendizaje conductistas nuevos y efectivos. El modelo de repetición de ejercicios era el método más usado, practicando la imitación y el refuerzo de la lengua oral y ninguna tolerancia hacia el error.

Lado (1964) resume una serie de principios en los que está basado el método audio lingual y que citaremos a continuación: hablar antes que escribir, frases básicas, modelos convertidos en hábitos, demostración del sistema de sonidos, control del vocabulario, enseñanza de los problemas, la escritura como representación del habla, modelos clasificados, práctica del lenguaje *vs.* traducción, estándar de lengua auténtica, práctica, formas de respuesta, velocidad y estilo, refuerzo inmediato, actitud hacia la cultura de estudio, contenido y finalmente, aprendizaje como resultado crucial. La crítica básica hacia este método fue que se convertía en una práctica mecánica de una lengua dada y que no animaba a los aprendices a crear un uso en aquello que aprendían.

### **2.7.1 Marco metodológico del CALL**

Pennington (1996) opina que los ordenadores ayudan a construir y elaborar las gramáticas de la L2 estructurando el *input* del alumno de diversas formas, por ejemplo, limitando y concentrando el lenguaje y las tareas que se realizarán. El ordenador es un método que responde positivamente y que ofrece a los usuarios una retroalimentación fiable y de alta calidad, generalmente de una manera instantánea. En su tratamiento del marco metodológico del CALL, indica cuatro aspectos a tener en cuenta:

- Jugador 1: alumno
- Jugador 2: promotor (diseña materiales)
- Jugador 3: evaluador (analiza el programa educativo de ordenador)
- Jugador 4: profesor

Una configuración adecuada de los métodos multimedia contribuirá a facilitar el conocimiento de la L2. Gimeno Sanz (2002), en relación al diseño y creación de paquetes multimedia CALL, considera que, una vez se ha tomado la decisión de crear algo, es esencial considerar los siguientes aspectos:

- El grupo de destino
- El nivel de lenguaje del grupo de destino
- El propósito del paquete en sí mismo: para uso general o específico
- El modelo metodológico
- El uso apropiado de la tecnología

Desarrollar CALL requiere una colaboración estrecha entre el profesor de lenguas y el programador. El diseñador gráfico es importante, y un ingeniero de sonido y un técnico de vídeo serán necesarios si el paquete ha de contener cantidades substanciales de sonido e imagen en movimiento.

Levy (1997) ha indicado que el papel del aprendizaje de lenguas asistido por ordenador se puede dividir en dos funciones principales: una como tutor y otra como herramienta. Su encuesta internacional sobre CALL en el año 1990 indicó algunos hallazgos relevantes, en los que la mayor parte de las personas que respondieron describieron ese tipo de aprendizaje como una herramienta. Asimismo, indicaron en ese año que resultaba un obstáculo el trabajar con materiales CALL en desarrollo, ya

que muchos carecían de validez académica y ese desarrollo de programas educativos por ordenador<sup>15</sup> contribuía a una promoción personal o a una prioridad institucional.

Phillips (1986) sugiere el análisis de siete facetas de materiales CALL:

1. Tipo de actividad (*juego, prueba o ejercicio*).
2. Tipo de aprendizaje: énfasis en el aprendizaje.
3. Foco de atención en el aprendiz: en lo que trabaja el alumno mientras interactúa con la lección.
4. Foco de atención en el programa: propósito lingüístico de la actividad.
5. Dificultad del nivel de lenguaje: nivel de lengua usado en el programa.
6. Dificultad del programa en sí: qué nivel de tolerancia tiene el programa con el alumno.
7. Gestión de la clase: cómo se usa el programa y se integra en el conjunto de la clase.

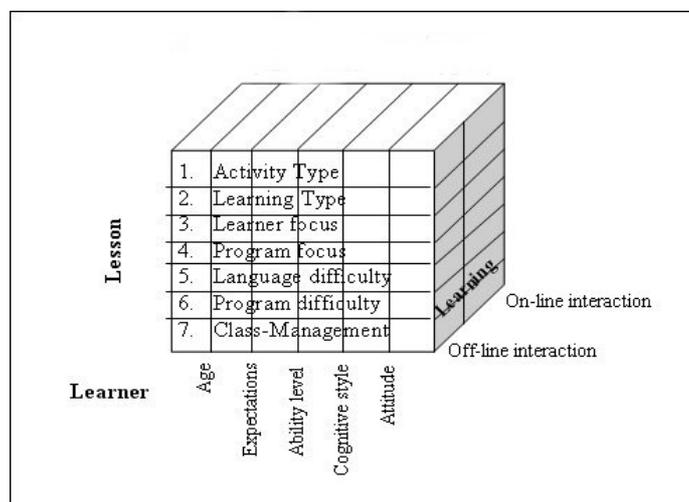


Figura 15. Chapelle y Jamieson, (1991:58). Dimensiones de análisis, alumnos y aprendizaje

<sup>15</sup> Levy también resalta que existen 24 disciplinas y campos de estudio en los cuales los desarrolladores de programas educativos por ordenador y los investigadores creen que han influenciado su trabajo.

Las dimensiones de análisis de lecciones, estudiantes y aprendizaje, según Chapelle y Jamieson (1991:58), quedan resumidas en la figura 15. Las características de los alumnos son importantes para asesorar la efectividad del CALL. Así pues, un programa educativo apropiado puede ser diferente para alumnos de diferentes edades, las expectativas basadas en experiencias de aprendizaje anterior serán importantes, el nivel de habilidad, el estilo cognitivo (independencia de campo / dependencia y reflexión / impulsividad), la relación dinámica entre la actitud y las variables del CALL contribuirán a asesorar su trabajo.

Estebanell *et al.* (2003) y Estebanell (1998) afirman que los alumnos suelen disfrutar si el CALL que usan tiene significado y es apropiado para ellos, e intenta aplicar dichos conceptos con alumnos de Magisterio como se traduce en el tipo de enseñanza de Nuevas Tecnologías aplicadas a la enseñanza. Según esta autora, debemos tratar de reflexionar profundamente acerca de las posibilidades y repercusiones de tipo didáctico de las TIC, generando un bucle permanente entre el conocimiento, el uso y la reflexión. Este proceso debe interiorizarse hasta convertirse en saber personal.

### **2.7.2 La entrada y salida de datos en los medios multimedia**

Brett (1999) argumenta que el subministro de la entrada de datos en L2 es un prerrequisito para SLA, es decir, la adquisición de una segunda lengua, y la tecnología multimedia proporciona una variedad de medios que preven esa entrada de datos: sonido, animación, texto escrito, vídeo y gráficos. El beneficio del multimedia es que puede proporcionar una entrada de datos lingüísticos que simultáneamente transmita la misma información a través de medios diferentes, ya que se dan canales diferentes de procesamiento de información: un canal acústico gracias al

sonido y un canal visual gracias a la animación del vídeo y del texto escrito.

El multimedia necesita proporcionar recursos extra que le permitan recuperar el significado desde la entrada de datos, aumentando este último punto gracias a la ayuda de los subtítulos, los diccionarios y las tareas de comprensión, con las cuales los alumnos tienen las herramientas necesarias para entender esa información. También, esta tecnología necesita que el alumno pueda tener el control para utilizar dichas características. El alumno puede ir a su ritmo gracias a la interacción personal con el ordenador.

Holliday (1999: 182) define la palabra *input* diciendo que en su sentido amplio se puede referir a todo el lenguaje de la lengua que se aprende, a la que el alumno está expuesto y que potencialmente le proporciona un conocimiento sobre ella. Los orígenes de investigación de la entrada de datos desde la perspectiva interaccionista se iniciaron con Ferguson (1971) cuando se refería al concepto de *foreign talk* o habla extranjera, como registro específico de *input* simplificado utilizado por los hablantes nativos cuando hablaban con aprendices de segundas lenguas.

Otros términos que necesitan definición son el *Comprehensible input* (Krashen 1977, 1980, 1985) aplicando el *Monitor Model Comprehensible Output* (Swain, 1985), quien indicó que la habilidad de los aprendices para entender las entradas de datos no garantizaba la adquisición de ciertas formas y estructuras, incluso cuando los aprendices que intentaban procesar el significado de un mensaje, podían no prestar mucha atención a la forma.

Debemos también considerar algunos factores internos en la adquisición de lenguas relacionados con los estudios que hemos citado en relación a la memoria.

Ellis (1997:119) describe el proceso del conocimiento implícito del aprendizaje (ver figura 16) según el cual en el primer estadio, el *input* (entrada de datos) se convierte en *intake* (internalización); éste implica que el aprendiz sea consciente de las características del lenguaje en el *input*, adsorbiendo las entradas de datos en la memoria a corto plazo y comparándolas a las producidas en el *output* o salida de datos.

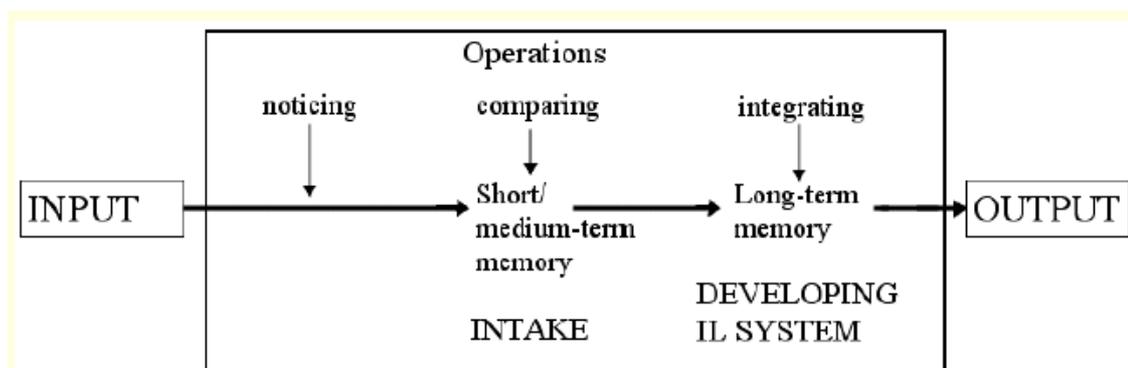


Figura 16. Ellis (1997:119). El proceso del conocimiento implícito del aprendizaje

Otros autores han dado otras definiciones para diferenciar el *input* y el *intake*:

*“Los datos de lenguaje potencialmente procesables y que están disponibles para el aprendiz de lenguas” [input]... “la parte del input que ha sido procesada y convertida en conocimiento de alguna forma [intake]”.*

Sharwood (1993: 167)

También Holliday (1999) cita estudios de Day y Shapson (1991) y Lightbown y Spada (1990) en los que también se ha sugerido que la adquisición de ciertas características lingüísticas puede requerir además *una atención hacia la forma* durante la retroalimentación en la producción de los aprendices.

Actualmente, en la investigación del *input* y del *output* se puede añadir el haber investigado los siguientes puntos:

- a) Los aprendices no están igualmente listos para todos los aspectos de la lengua de estudio al mismo tiempo.
- b) El desarrollo ocurre como resultado de los diferentes niveles de atención en el *input*.
- c) La retroalimentación correctiva en el *output* tiene un rol en el aprendizaje de lenguas, pero no es universalmente la mejor estrategia.

Todos los puntos tratados en relación a las entradas y salidas de datos son aplicables al CALL, ya que los ordenadores influyen el ambiente en el que los aprendices están expuestos a la segunda lengua. Pero no hay muchos programas que presten atención a la interacción en los CD-Roms y programas de *software*. De hecho, muchos de estos programas no interactúan en absoluto y todas las respuestas que puede dar un ordenador están basadas en bancos de datos pre-programados y en audio y vídeo pre-grabado.

Aunque el ordenador no entiende el lenguaje natural, es capaz de recibir (*input*) y producir (*output*) múltiples cantidades de datos. Sin embargo, para que este aspecto sea más viable, debe tener la capacidad de interactuar con el usuario aproximándose a la comunicación real, si bien existen aspectos no directamente relacionados con la mejora en la enseñanza de lenguas.<sup>16</sup>

Brett (1999) comenta que el uso de la entrada de datos bimodal (canal visual y auditivo) puede facilitar y permitir que los alumnos trabajen la

---

<sup>16</sup> Meredith (1978) llevó a cabo un estudio que asesoraba la impulsividad / reflexión de los alumnos y encontró que los alumnos impulsivos realizaban mejor una frase oral (en español) cuando el programa CALL les forzaba a esperar antes de responder.

fonología. También señala que hay cuatro caminos en los cuales el multimedia puede ayudar a trabajar las entradas de datos:

- La previsión de tareas en pantalla designadas para ofrecer especial atención a características lingüísticas.
- La existencia de una retroalimentación<sup>17</sup> para estas tareas.
- Algo que acabemos de resaltar o pronunciar.
- El abastecimiento de recursos que los alumnos pueden usar para negociar la comprensión.

Los medios multimedia necesitan repetir las condiciones que se cree que desarrollan el control sobre el conocimiento de la L2.

## 2.8 Los laboratorios de idiomas

En el primer capítulo, al referirnos a los aprendizajes conductistas y el modelo de repetición de ejercicios, ya se ha introducido muy brevemente qué tipo de herramienta de uso era el laboratorio de idiomas y también se ha introducido el laboratorio portátil ofrecido por el *Departament d'Educació*. En las sucesivas líneas vamos a ver las diferentes posibilidades que ofrece esta herramienta desde un punto de vista global. Para ello, es necesario primero ver la evolución de esta maquinaria, hacer una referencia a qué países la utilizan de una manera más regular y con más avances. Posteriormente, vamos a distinguir entre los laboratorios antiguos, que empleaban un sistema analógico, y los laboratorios actuales, que son digitales.

---

<sup>17</sup> Brandl (1995) investigó las preferencias en el tipo de *feedback* en un programa alemán de CALL, indicando que los tipos de *feedback* eran: Respuesta correcta o incorrecta (1), Descripción gramatical de la respuesta correcta (2) y localización del error y respuesta correcta (3). Los alumnos de nivel alto optaron por el primer y tercer tipo, mientras que los alumnos de nivel bajo preferían la descripción gramatical correcta y visualizar la respuesta correcta.

### **2.8.1 Los inicios de los laboratorios de idiomas**

Según Davies *et al.* (2005), la tecnología digital se remonta a la invención del ordenador digital en Manchester en 1948, si bien hasta mediados de la década de los años 70 no aparecen los primeros micro ordenadores a precios razonables. La inclusión del multimedia dentro del laboratorio digital multiplica las diversas formas de difusión de contenidos, siendo éstas la radio, la televisión digital o terrestre, la grabadora de sonidos, el reproductor de CD o DVD, la Internet, el teléfono móvil o las agendas electrónicas (PDA, siglas en inglés de *Personal digital assistant*). El denominador común en estos aparatos es que la mayoría pueden reproducir vídeo - al cual se refiere Hill (1999) diciendo que es posible estudiar el movimiento de los labios, los gestos relativos propios del lenguaje oral- y por supuesto audio.

Davies *et al.* (2005) se refieren a los laboratorios de idiomas diciendo que fueron populares en escuelas de educación secundaria y otras instituciones a finales de la década de los años 60 y principios de los 70 del siglo anterior. Fueron inicialmente percibidos como una solución al problema del aprendizaje de una lengua a corto plazo para un número elevado de alumnos y se consideró que el invento valió la pena. Sin embargo, su uso decreció a finales de la década de 1970, en parte debido a la falta de confianza en la tecnología, pero también porque se asoció estrechamente con un modelo conductista pasado de moda para la enseñanza y aprendizaje de lenguas. Las razones por las cuales no se obtuvieron los resultados esperados fueron: la falta de un entrenamiento adecuado de los profesores que deseaban emplearlo y la falta de creatividad en proponer actividades más que ejercicios mecánicos; dos aspectos que, en la práctica, aún han de mejorarse en la actualidad.

Stern (1992:10) remarca que la tecnología se convirtió en un aspecto central del nuevo método audio lingual y el laboratorio formaba parte de una nueva era. Igualmente, según Underhill (1989), refiriéndose a esta

herramienta y refiriéndonos a la mejora de la pronunciación, las máquinas probaron ser el objetivo primordial para ayudar al profesor mediante un número inacabable de ejercicios, un modelo de pronunciación perfecto y una manera de liberar al profesor de actividades intelectuales en la clase. Rivers (1990:274), al referirse a la desaparición del método audio lingual y a la potenciación del método comunicativo, comenta que su uso pareció irrelevante y aburrido.

Celce-Murcia *et al.* (1996:311) se refieren a la cinta de casete y la grabadora como herramienta dentro del laboratorio de idiomas, indicando que fueron el medio y el lugar adecuado para que se produjera ese aprendizaje. Con la desmitificación de la psicología conductista y el desencanto resultante en el campo de la lingüística con el método audio lingual, el marco de los laboratorios de idiomas y la tecnología instruccional cayó en desgracia. De hecho, esto ocurrió no tanto porque la tecnología en sí tuviera la culpa, sino porque los métodos y los materiales estaban pasados de moda, tenían limitaciones de aplicación y eran aburridos o carentes de estímulo. Dichos autores comentan que el renacimiento del laboratorio de lenguas representa un triunfo de la tecnología sobre el método y proporciona un testimonio al papel que la tecnología instruccional puede jugar en la enseñanza de lenguas.

A mediados de la década de 1980, el laboratorio volvió a tener un nuevo sentido, gracias en parte a las mejoras tecnológicas, pero también debido a unas mejoras en las herramientas de control, al hacerlas más manejables y sencillas, unos materiales más imaginativos y un mejor diseño del propio laboratorio. Al mismo tiempo, el autoaprendizaje se puso de moda y se produjo una riqueza de nuevas ideas tales como el uso del trabajo en pareja, el empleo de técnicas de situación (*roleplays*), juegos comunicativos, etc., descritos por Ely (1984).

## 2.8.2 La evolución hasta la fecha actual

Hoy en día, las grabaciones de audio y el marco del laboratorio de idiomas continúan siendo valorados para la práctica de la pronunciación, particularmente debido a la aparición en el mercado comercial de programas auxiliares.

Murphy (1993), al referirse a los laboratorios de idiomas, utiliza el término LARC para designar el *Language Acquisition Resource Center*, en una cita que engloba información interesante relativa a la correcta utilización de este recurso informático y a su uso comunicativo en entornos de aprendizaje:

*"If used wisely, a language laboratory is an example of a modern technological resource that can provide a rich source of learning opportunities for ESL learners. Such opportunities can be provided within whole-class instructional settings and through activities that aim to facilitate peer collaborations as well as self-directed learning. However, it will probably fall to L2 classroom teachers, working in cooperation with LARC personnel and program administrators, to guide ESL students in the use of audio and video material, computer programs, and other technological resources through procedures reflective of the current paradigm for L2 instruction".*

Murphy (1993:14)

Dada esta disponibilidad de material auténtico, el laboratorio de lenguas puede proporcionar acceso casi ilimitado al discurso nativo y a todas sus variantes (dialectos, registros, sexo, etc.). Además, el laboratorio puede proporcionar a los alumnos otras condiciones deseables, como son una mayor práctica que no es posible en el ambiente típico de clase, un ambiente desinhibido para imitar sonidos, la posibilidad de comparar su voz con la del modelo y el control de esos materiales, de sus secuencias y del ritmo.

Ya en la década de los años 90, Beauvois (1996) realizó un estudio en el cual se comparaba la producción del alumno en dos ambientes y

concluyó que la conversación generada en el laboratorio de lenguas fue superior al de los debates de clase sobre los mismos tópicos, tanto en calidad como en cantidad.

### **2.8.3 El mercado de laboratorios americano**

Actualmente, en el mercado de laboratorios portátiles existen diversos productos que desafortunadamente suelen ser sólo accesibles, por su coste, en América del Norte (Estados Unidos y Canadá), países pioneros en este tipo de tecnología.

A tenor de los diferentes productos analizados, podemos afirmar que suele existir un denominador común de recursos: una consola para el profesor con la que se ofrecen sesiones de entrenamiento a los profesores que se encargan de utilizar estos productos, un ordenador ‘servidor’ en el que los recursos están almacenados y se ofrecen a los alumnos y auriculares con micrófonos incorporados (normalmente a prueba de malas manos, es decir, diseñados para resistir y ser utilizados muchas horas). Algunas empresas como *CAN 8 Virtual Lab* empezaron la distribución de estos productos en Canadá en 1967, contando también con la ayuda del *National Research Council of Canada* (NRC). *Sanako*, una empresa americana multinacional en colaboración actualmente con la República Popular China y Finlandia, introdujo el primer laboratorio en el año 1961<sup>18</sup>.

---

<sup>18</sup> [documento en línea] <http://www.sanako.com>



Figura 17. Laboratorio de idiomas en EE.UU. Sistema de ahorro de espacio y almacenaje en techo

Algunos productos norteamericanos (por ejemplo *Educational Media, LLC*) apuestan por la reducción o adaptación de espacio e introducen cajas plastificadas colgadas en el techo del aula (ver figura 17). Mediante un mecanismo, bajan a una altura estándar donde el alumno simplemente se ha de colocar los auriculares, activar un micrófono y ponerse a trabajar, en este caso en formato papel. Es el profesor el que, mediante una consola, interactúa con los alumnos. Éstos no manipulan ninguna máquina, con lo cual el riesgo de deterioro disminuye.

#### **2.8.4 El sistema analógico y el sistema digital**

Renschel (2004) comenta que en 1999, un 10% de los laboratorios vendidos en EEUU eran digitales, en el 2002 ya fueron un 85%, y un 99% en 2004. A finales de los años 90 y a principios del 2000 se introdujeron los laboratorios híbridos, de transmisión analógica y almacenamiento digital.

Las razones por las cuales se empleaba un sistema analógico fueron la existencia de una amplia biblioteca de materiales en dicho formato (cilindros, discos, cinta magnética), la limitación de fondos, la saturación de reproductores de casete y, finalmente, la tecnofobia o miedo al uso de tecnología más avanzada.

El hecho de utilizar los productos digitales -entiéndase discos duros, CD-Roms, memorias portátiles- supone la irrupción tecnológica por la cual no sólo se trabaja con audio, sino con medios multimedia. ¿Qué la hace diferente? Existen varios factores: la habilidad de poder realizar la grabación comparativa, la posibilidad de reproducir la pista del programa mientras simultáneamente graba al alumno, así como el permitir al estudiante trabajar a su propio ritmo. Ahora bien, el próximo apartado nos lleva a considerar cuál es el producto más interesante a tener en cuenta.

## **2.9 Tipos de laboratorios**

El laboratorio digital está ligado al uso adecuado de *software* que proporcione la integración de vídeo, un procesador de palabras y aplicaciones informáticas. Se pueden distinguir dos tipos: los que combinan sólo *software* y los llamados híbridos.

El primer tipo es el típico ordenador que solemos tener en casa con la diferencia de que existe una comunicación con el profesor utilizando un controlador especial, mientras que el segundo tipo ya requiere un cableado mucho más profesional e interfaces para combinar la ayuda de la comunicación verbal y las señales de control. Conceptos tales como la versatilidad, la facilidad de movimiento entre diversas aplicaciones, la interactividad, el potencial para la intervención del profesor y el aprendizaje independiente son características básicas en los laboratorios digitales.

En la actualidad, el uso de los laboratorios, sean o no portátiles, depende de varios factores:

- 1) De la prioridad de uso del profesor, si lo cree conveniente.
- 2) De la existencia de aulas de idioma que faciliten el uso de esa tecnología, ya que la mayoría de docentes han de cambiar de aula y no existen muchas salas destinadas al aprendizaje de una LE.
- 3) La utilidad que se le pueda dar a la máquina, así como su máximo rendimiento, no sólo como reproductor, sino como complemento de grabación, uso de materiales exclusivos, uso del micrófono y de los auriculares.

Un componente importante como la motivación puede tener una incidencia distinta dependiendo de su uso. Hunter (2001b), ver figura 18, en la siguiente página, considera que el control del uso por parte del alumno hace éste tenga un alto nivel de motivación, mientras que, el que lo dirija el profesor y que se vincule el uso remedial como mejora de varios aspectos del habla, hace decrecer dicho factor. La parte inferior del gráfico de la siguiente figura explica la diferencia entre el laboratorio CALL y el laboratorio para uso autodidáctico basado en CALL. Las diferencias principales residen en que, si es el alumno quien controla su uso y su mejora, los programas educativos pueden ser muchos, así como su ámbito de uso: auto-mejora, trabajo remedial y deberes.

Para centrar la cuestión, conviene mencionar los aspectos que Hunter incluye en el uso del laboratorio: todo lo relacionado con texto e hipertexto, imagen, sonido, multimedia, vídeo y animación, juegos y funciones de búsqueda en Internet; elementos que no totalmente están relacionados con el uso que intentamos comentar aquí, es decir, el de la integración de una interacción entre la máquina y el alumno, con audiciones y repeticiones de producciones orales.

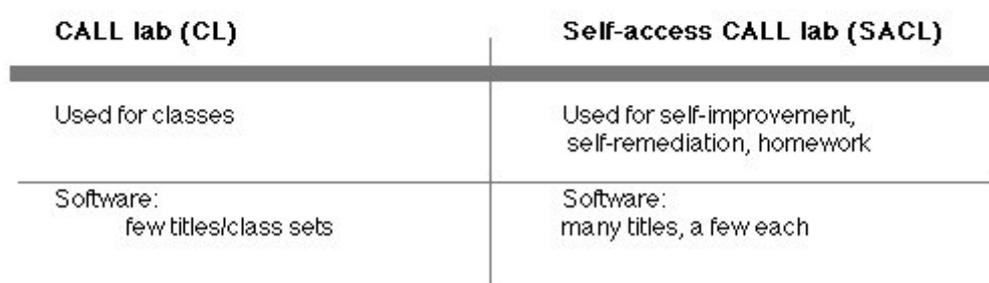


Figura 18. Hunter (2001b). La motivación y el uso de los laboratorios <http://www.info.kochi-tech.ac.jp/lawrie/CALLL/index.html>

Naiman (1992), al referirse al laboratorio de lenguas, incide en que no se necesita un laboratorio caro para enseñar pronunciación, ya que el hecho de que los alumnos puedan grabarse en cintas para reproducir y controlar sus producciones orales en clase así como grabar conversaciones naturales hace que ninguna de estas actividades requiera un laboratorio. De todas maneras, las siguientes páginas van a mostrar ejemplos de tecnología que pueden llevar a cabo una amplísima variedad de interacción entre alumnos.

Gimeno Sanz (2002) apuesta por utilizar actividades del laboratorio de idiomas con alumnos para registrar sus voces y compararlas con un modelo pre establecido, que abarque desde palabras aisladas y frases a unidades más largas de significado, como intercambios en un diálogo.

Remschel (2004) se refiere a los cuatro principios que se consideran indispensables a la hora de referirse a la tecnología digital.

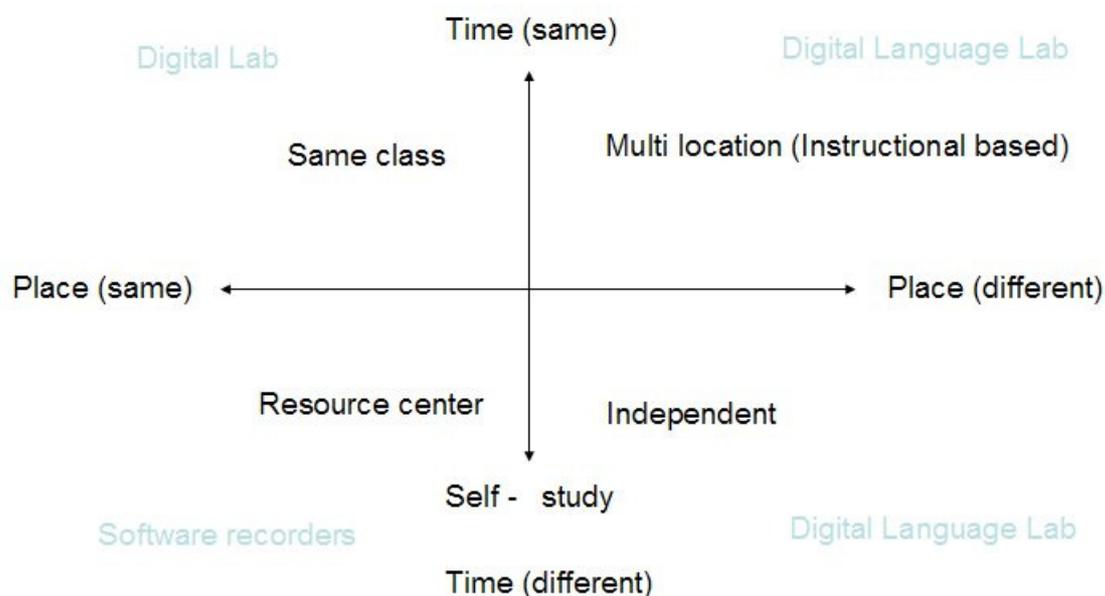


Figura 19. Remschel (2004). “Digital Technology: where we are, where we are going”

En la figura 19, Remschel (2004) divide el uso de la tecnología refiriéndose a su utilización en un lugar u otro, a la vez que sea de un producto u otro, o en un momento diferente. Según estas variables, los productos se pueden englobar para ser utilizados en el aula de idiomas tradicional, en el centro de recursos, en un lugar independiente o en diferentes espacios. El uso del laboratorio digital permite trabajar a la vez o en un momento diferente e inclina la balanza hacia el estudio personal.

## 2.10 El reconocimiento automático de voz. Análisis teórico

En CALL, la señal analógica se convierte en digital y ésta se refleja en una secuencia lógica de declaraciones, expresadas con los números binarios 0 y 1. El desarrollo de tecnología sofisticada, como el ordenador digital, ha introducido la creación de programas de *software* capaces de

extraer el tono, la frecuencia y la intensidad del habla grabada de una manera rápida y precisa, así como la visualización de los modelos acústicos y espectrográficos en una pantalla de un ordenador. El proceso de digitalización de una señal analógica para que pueda ser usada con *software* se resume de la siguiente manera:

- 1) El habla se graba vía micrófono en una cinta de forma analógica.
- 2) Ésta se convierte en digital con un convertidor del ordenador.
- 3) El ordenador la procesa.
- 4) Los resultados se pueden enviar a un monitor, impresora, etc.

La síntesis fonémica es un proceso por el cual cada fonema, o sonido, de un lenguaje está representado por un código digital o filas de símbolos. Este código se traduce por el sintonizador en modelos particulares de frecuencia, duración y amplitud, que modelan los sonidos del habla.

Ehsani y Knodt (1998) revisaron el estado del reconocimiento de voz y de la tecnología con unas aplicaciones en concreto y sugirieron cómo crear entornos de aprendizaje robustos. Según ellos, el CALL ha emergido como una alternativa tentadora a las maneras tradicionales de suplementar o reemplazar la interacción directa profesor-alumno, con los laboratorios de idiomas y el estudio personal basado en cintas de audio. Dichos autores relacionan diversa literatura en relación al aprendizaje asistido por ordenador y llegan a la conclusión de que falta un marco teórico unificado para el diseño y la evaluación de los sistemas con CALL (Chapelle 1997, Hubbard 1988), así como una ausencia de evidencia empírica conclusiva para los beneficios pedagógicos del aprendizaje de lenguas con ordenadores (Chapelle 1997, Dunkel 1991 y

Salaberry 1996). La siguiente cita resume los aspectos explicados anteriormente:

*“Los humanos y las máquinas procesan la información de una manera diferente fundamentalmente (Bernstein y Franco, 1996). Los procesos cognitivos complejos consideran la habilidad humana de asociar señales acústicas con significados e intenciones. Para un ordenador, por otra parte, el habla es esencialmente una serie de valores digitales. Sin embargo, a pesar de esas diferencias, el problema clave del reconocimiento del habla es el mismo para los humanos y para las máquinas: concretamente el encontrar la mejor equivalencia entre una muestra de sonido dado y la correspondiente palabra”.*

Ehsani y Knodt (1998:47)

### **2.10.1 Origen y definiciones**

O’Shaughnessy define la comunicación oral de la siguiente manera:

*“The transfer of information from one person to another via speech, which consists of variations in pressure coming from the mouth of a speaker”.*

O’Shaughnessy (1987:1)

Faudez (2000) comenta los fundamentos del tratamiento del habla y cita:

*“La generación del habla por parte de las personas, consiste en la creación de una onda de presión sonora que se propaga a través del aire a una velocidad de unos 340 metros por segundo. La concatenación de unos determinados sonidos y en un orden prefijado, característicos de cada idioma, constituyen el mensaje”.*

Faudez (2000:32)

La investigación moderna del habla se inició en los años 1930, cuando se desarrolló un método de transmisión práctica digital denominado PCM (*Pulse Code Modulation*). La invención del espectrógrafo en 1946 amplió mucho la capacidad de análisis del habla. No fue hasta los tempranos

años 70 del siglo pasado cuando el reconocimiento del habla explotó su trabajo.

### **2.10.2 Las técnicas del reconocimiento de habla**

Por otra parte, Llisterri (2001:16) define las técnicas de reconocimiento de habla como una comparación, en los métodos más sencillos, entre los patrones acústicos previamente almacenados en el sistema y los enunciados producidos por el hablante.

Varios autores como Stevens (1998), Ladefoged (1996) y Kent y Read (1992) se han referido al análisis de los términos comentados a continuación. En este punto, conviene tener clara una nomenclatura y unos términos que son necesarios para profundizar en los aspectos tratados en esta tesis. También son de gran ayuda las representaciones visuales y comentarios aparecidos en Llisterri (2006) acerca de elementos que veremos a continuación.

## **2.11 Elementos importantes al analizar el habla**

Existen una serie de elementos que conviene mencionar y analizar su funcionamiento para un mejor conocimiento de cómo trabaja el reconocimiento de voz.

### **2.11.1 El espectrograma**

Uno de los elementos esenciales para analizar el habla es el espectrograma -también llamado monograma- que convierte una forma de ondas del habla en dos dimensiones (amplitud y tiempo) en un modelo tri-dimensional (amplitud, frecuencia y tiempo). Con el tiempo y la

frecuencia en el análisis horizontal y vertical respectivamente, la amplitud se mide por la oscuridad del visualizador. Los sonidos sonoros son los que se marcan en el espectrograma, debido a un incremento de la amplitud del habla cada vez que la vocal se acerca. Estos sonidos sonoros son las vocales (incluidos diptongos), excepto cuando se susurran. Los sonidos sordos causan unos modelos rectangulares oscuros puntuados con puntos claros debido a las variaciones instantáneas de energía.

Komissarchik *et al.* (2000) comentan que todas las representaciones reciben el nombre de espectrogramas, o representaciones de la amplitud en función de la frecuencia. La señal de voz es limitada en banda, a unos 8 kHz. Sin embargo, la mayor parte de la información se encuentra en los primeros 4 kHz, que es aproximadamente el ancho de banda utilizado en las comunicaciones por vía telefónica. En relación a la prosodia, el habla conversacional emplea 150-250 palabras por minuto, incluyendo pausas, que promedian unos 650 ms (milisegundos) cada una. Por poner algunos ejemplos, la duración típica de una sílaba es de 200 ms, con vocales acentuadas, éstas promedian 130 ms y la duración de otros elementos fónicos alrededor de 70 ms. Así pues, la vocal /ə/ (schwa) es típicamente de 45 ms, mientras que los diptongos en conversación promedian 180 ms. El promedio vocálico es 30 ms más largo que las consonantes sonorantes, que son 5 ms más largas que las obstruentes. Los diptongos son 75 ms más largos que las vocales. Mediante un micrófono es posible captar la onda de presión sonora emitida por un orador y convertirla en señal eléctrica, digitalizada y poder tratarla en un ordenador.

Lambacher (1996b), al definir las propiedades de los sonidos, se refiere a la frecuencia de un sonido diciendo que éste está determinado por la variación de la presión de aire o el número de ciclos completados por segundo de una onda sonora (medidos en hercios) y, dependiendo de los sonidos, éstos tienen diferentes concentraciones de energía. Por poner

un ejemplo, el sonido fricativo /s/ en inglés, tiene una concentración de energía que oscila entre 4000 y 7000 hz, mientras que el sonido nasal /m/ tiene una baja concentración de energía, entre 250 y 300 hz.

Otro aspecto destacable de los espectrogramas correspondientes a fonemas sonoros es la presencia de zonas enfatizadas, o resonancias, y zonas desenfáticas, o también llamadas antirresonancias. Estas zonas se deben a la configuración física de las diferentes cavidades del tracto configurado por el sistema fonador<sup>19</sup>.

Los formantes aparecen como franjas horizontales, mientras que los valores de amplitud en función de la frecuencia se representan en tonalidades de grises en sentido vertical. Existen dos tipos de señales sonoras que se pueden visualizar: la de banda ancha, en la cual se ven los picos de voz, y la de banda estrecha, en la que se ven unas impresiones de marcas. Dichas señales dependen del ancho de banda del filtro que se haya utilizado para realizar el análisis frecuencial.

### **2.11.2 El osciloscopio**

Un osciloscopio mide la evolución temporal de la amplitud del sonido. Con él es posible obtener una imagen que representa la amplitud de la señal en función del tiempo, que suele denominarse oscilograma. Es importante destacar que tanto el micrófono usado como la sala en la que se realiza la grabación pueden afectar a las tasas de reconocimiento. Llisterri (2001:6) ha tratado de comentar los diferentes datos que se observan en las producciones del alumno en un osciloscopio, así como también la producción nativa o del profesor.

---

<sup>19</sup> Faudez (2000) nombra como componentes principales del sistema fonador humano a los pulmones, la tráquea, la laringe, la cavidad de la faringe, la cavidad oral y bucal y la cavidad nasal. El tracto vocal empieza a la salida de la laringe y termina a la entrada de los labios. El tracto nasal empieza en el paladar y termina en los orificios nasales. Los parámetros principales del sistema articulatorio son las cuerdas vocales, el paladar, la lengua, los dientes, los labios y las mandíbulas.

Más adelante, al analizar diversos programas de pronunciación, podremos profundizar en este aspecto.

Petek (1999) hace referencia a la necesidad de una gran fiabilidad en el interfaz de reconocimiento de voz:

*“In general, multimedia courseware can be considered as an efficient complement to the traditional ways of teaching. Interactivity of the courseware can balance between the teacher guided and open ended learning environments. It can potentially create and stimulate the learners’ motivation and interest. Therefore, it is important that the interactive multimedia courseware offers an efficient user interface”.*

Petek (1999, documento *online*)

### **2.11.3 Otros términos y consideraciones importantes**

Utilizando los principios que contribuyen al éxito en el entrenamiento de la pronunciación, Kenworthy (1987) presenta una lista de cinco principios que se han de aplicar en el entrenamiento de la ASR (*Automatic Speech Recognition*) o reconocimiento automático del habla, divididos en cuatro factores que se refieren al ambiente externo:

- a) Deben producir grandes cantidades de frases por sí mismos. La tutorización individual con instrucción humana suele ser demasiado costosa e impráctica. El tener que compartir la atención del profesor con otros reduce la cantidad de tiempo que cada alumno emplea en crear producciones en lengua extranjera.
- b) Deben recibir la corrección oportuna, para ello los profesores deben asegurarse de que hayan enseñado a los alumnos a producir – articular el sonido correctamente, más que a producir pares mínimos (*minimal pairs*).

- c) Deben oír diferentes modelos nativos, gracias al multimedia ello es posible.
- d) Se debe enfatizar la prosodia (amplitud, duración y tono).

Y uno al ambiente interno:

- e) Deben estar relajados en la situación de aprendizaje. Cuando se fuerza al alumno a producir sonidos que no existen en su lengua nativa delante de sus compañeros, tiende a perder auto confianza.

La ASR se puede usar en la enseñanza de la pronunciación en L2 como complemento del profesor humano. La máquina se hace cargo de esos aspectos de pronunciación práctica que corresponden a los principios anteriormente explicados.

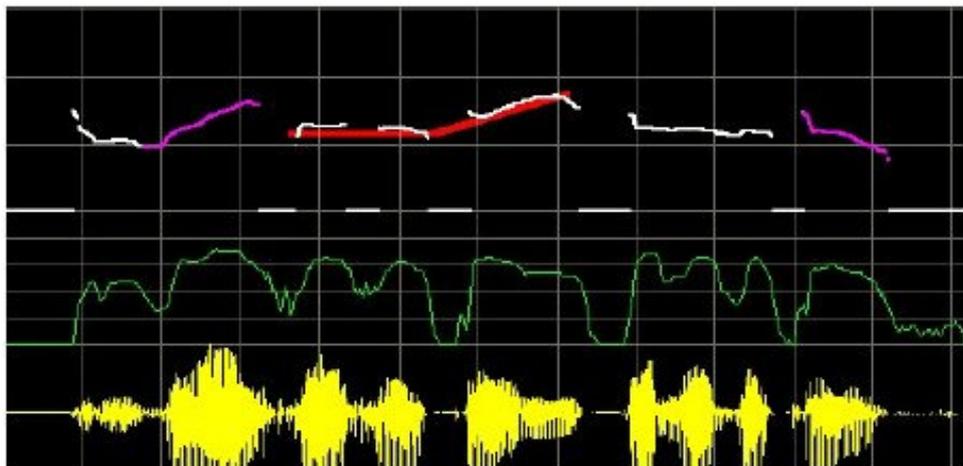


Figura 20. Llisterri (2001:6). Oscilograma, curva de intensidad y curva melódica con el programa WinPitch LTL (Germain y Martín, 2000, figura 2)

Los programas de reconocimiento de voz tratan de llevar a cabo los principios anteriores, algunos con mayor acierto que otros. En dichos mecanismos tratados en los dos puntos anteriores es posible observar la

llamada curva melódica (que se puede observar en la parte superior de la figura 20), es decir, la variación temporal de la frecuencia fundamental y la curva de intensidad (parte central de la misma figura 20) o evolución de la intensidad en el tiempo. Esta figura nos muestra dichos fenómenos así como un oscilograma (parte inferior del gráfico).

La figura 21, a continuación, nos muestra un dibujo que corresponde a un fragmento de un diptongo pronunciado por un hablante masculino.

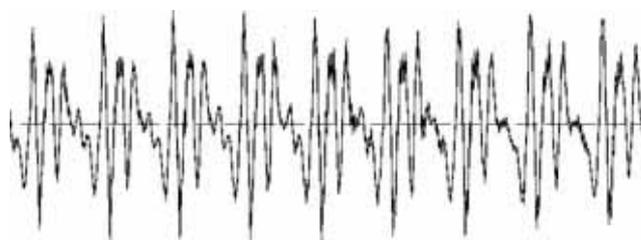


Figura 21. Komissarchik *et al.* (2000:2).Diptongo 'aw' dentro de la frase "How are you?".  
Pronunciación masculina

Komissarchik *et al.* (2000) comenta, en referencia a la anterior figura, que los topes representan la quema de energía que se produce después de que las cuerdas vocales se sacudan y el aire pase a través de ellas. Estos topes se llaman también puntos de excitación. Si una persona produce una excitación durante un segundo, decimos que su tono es de 100 Hz (herzios). El alcance de tono para hombres es de 80-120Hz, para mujeres de 180-220 Hz, y para niños puede llegar a ser de 400-500Hz<sup>20</sup>.

Pellegrino *et al.* (2001) nos ofrece una página web llamada *Ciberphone* a un nivel muy visual y auditivo que puede servir para mejorar la comprensión de ciertos aspectos técnicos descritos aquí de manera más breve.

<sup>20</sup> Si bien Faudez (2000) tiene unos parámetros superiores, diciendo que el margen habitual para locutores masculinos adultos del valor del tono (*pitch*) es de 50 a 250 Hz (período de 20 a 4 ms o 160 a 32 muestras al trabajar con frecuencias de muestreo  $F_m = 8$  kHz, mientras que para locutoras se encuentra entre 120 y 500 Hz (período de 8,3 a 2 ms o 66 a 16 muestras con  $F_m = 8$  kHz).

Dicha página web dividida en tres partes: una introducción a la naturaleza de los sonidos y a su propagación, la descripción de las ondas sonoras y las tipologías de sonidos, y finalmente, las nociones espectrales y los espectrogramas.

En la figura 22, a continuación, que se refiere a un conversor de texto a voz, Faudez (2000) afirma que en los sistemas de conversión texto-voz no es totalmente necesario obtener un excelente grado de inteligibilidad segmental. Según él, la inteligibilidad en la ASR está relacionada con la facilidad para comprender la señal oral. Normalmente, se acostumbra a relacionar la inteligibilidad con el proceso segmental. La calidad es un indicador de la “naturalidad” de los sonidos.

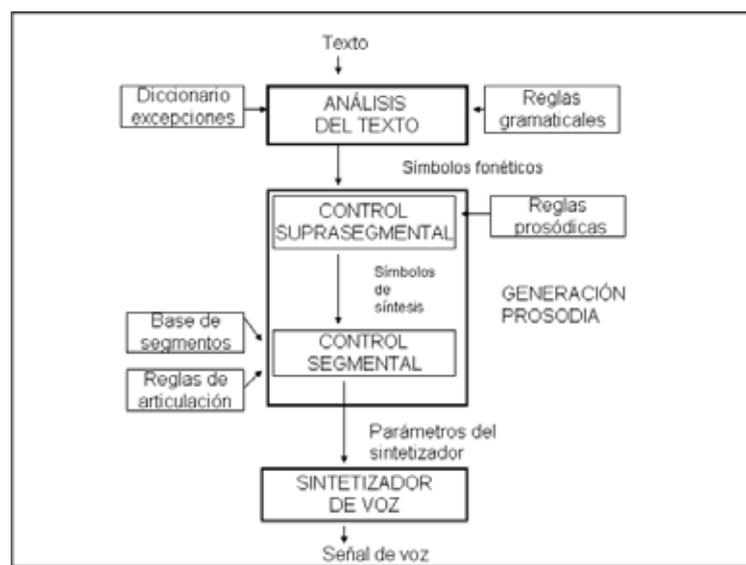


Figura 22. Faudez (2000:82). Conversor de texto a voz

Por lo que hace referencia al procesado prosódico o suprasegmental (conjunto de factores que afectan la pronunciación de manera global, como la entonación, el ritmo y la intensidad del habla), si el texto es relativamente largo, es muy importante que la voz sintetizada tenga una buena entonación y ritmo, y que sea agradable escucharla. Se puede considerar que un convertidor texto-voz consta de tres grandes bloques:

- Análisis lingüístico del texto
- Generación de la prosodia
- Síntesis de la señal de voz

La facilidad para realizar la transcripción fonética depende del idioma que se está tratando. Por ejemplo, resultará más sencillo en español que en inglés, ya que en el primer caso existe una gran concordancia entre la grafía y los fonemas, la cual simplifica el análisis. En cambio, en el inglés, hay que recurrir a un análisis morfológico y existe una complejidad de reglas que ha originado otras alternativas como los modelos conexionistas (estándares fonéticos europeos con caracteres ASCII).

El análisis de la estructura sintáctica permite a los módulos posteriores de generación prosódica determinar correctamente la localización de los grupos de entonación y, para resolver las ambigüedades en la transcripción de los homónimos (es decir, las palabras que en inglés se escriben igual, pero pueden tener pronunciaciones distintas), se hace necesario un análisis semántico del texto que ayude a determinar la pronunciación correcta.

## **2.12 Los diferentes tipos de reconocimiento de voz**

Según Wachowicz *et al.* (1999:255), el CALL de habla interactiva utiliza una variedad de reconocedores del habla y distingue cuatro tipos:

- a) Los sistemas de ASR de hablante independiente, que intentan reconocer casi cualquier hablante de inglés sin tener que entrenar el reconocimiento de voz en el tono o en el modelo de habla de un hablante individual. Estos sistemas son más apropiados en

aplicaciones de aprendizaje de lenguas, especialmente para hablantes no nativos. Estos sistemas han sido diseñados para usar varios tipos de voz y acentos regionales, así que se supone que aceptan todo tipo de hablantes nativos así como entrada de datos comprensibles y no-nativos.

- b) Los sistemas ASR de hablante dependiente, en que el hablante ha de entrenar su voz, y son adecuados para dar órdenes y poder utilizar una computadora.
- c) Los sistemas de ASR de voz discreta, cuyo propósito es reconocer palabras individuales y cuya entrada de datos se ha de separar por pausas. No serían adecuados para tareas a nivel frase.
- d) Los sistemas de ASR continua, que sirven para procesar frases más largas, tipo *IBM Via Voice* o *Dragon's Naturally Speaking*.

Como regla general, un reconocedor automático del habla no puede procesar correctamente el habla que difiere de la manera con la que el habla ha sido diseñada. Esta es una de las razones por las que los sistemas comerciales de dictado, cuando han sido pensados para inglés americano estándar, tienen una actuación pobre cuando encuentran muestras de habla acentuada tanto por hablantes no nativos como por hablantes de otros dialectos.

LaRocca (1999) también exploró las posibilidades del reconocimiento del habla e indicó que desde 1994 ha sido posible trabajar con esta tecnología con un coste menor. El uso de la ASR trajo nuevos beneficios al CALL en general y al autoaprendizaje en general. Los alumnos pueden practicar la lengua oral con una retroalimentación inmediata. De todas maneras, hace hincapié en trabajar, especialmente a nivel principiante, con reconocedores de voz híbridos, que procesan señales avanzadas juntamente con enlaces medidos de manera tradicional. Según este

autor, es de vital importancia el que cada lección tenga su propio vocabulario para producir reconocimiento de voz y esas producciones deben ser cortas. La forma en la que LaRocca utiliza el reconocimiento de voz es diversa:

- a) La forma más común es cambiar el teclado y el ratón por el micrófono con el paradigma de las respuestas múltiples. Según este paradigma, el programa hace escoger al alumno entre tres o cuatro alternativas por respuesta, entre las cuales el aprendiz ha de leerla en voz alta. El sistema está entrenado para reconocer esas alternativas.
- b) Los alumnos están obligados o incitados mediante texto, vídeo o audio a realizar el ejercicio.
- c) Para trabajar el análisis contrastivo de la pronunciación, se presentan dos palabras para trabajar los pares mínimos.

En relación al entrenamiento de la pronunciación, los tutores de pronunciación con voz interactiva hacen repetir a los alumnos palabras y frases. La clave para que ello sea un éxito es la retroalimentación correctiva, más específicamente un tipo de retroalimentación que no confie en la propia percepción del alumno. Mientras que la gramática y el vocabulario de un tutor de pronunciación son comparativamente simples, el procesamiento del habla tiende a ser complejo y se debe adaptar para reconocer y evaluar el habla del alumno, que no siempre es fluida. Un tutor de pronunciación, como contraste, debe entrenarse en reconocer y corregir las desviaciones de pronunciaciones nativas comunes.

Ehsani y Knodt (1998) citan a Berstein (1997), Franco *et al.* (1997), Kim *et al.* (1997) y Witt y Young (1997), que tienen diversas técnicas para reconocer de manera automática y anotar el habla no nativa. En

términos generales el procedimiento consiste en construir modelos de pronunciación nativa y entonces medir las respuestas no nativas con los modelos nativos. Ello requiere modelos entrenados tanto en respuestas con datos de habla nativas como no nativas en la lengua de estudio y suplementadas por una serie de algoritmos –un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema- para medir las variables acústicas que se han probado útiles en distinguir el habla nativa de la no nativa. Estas variables incluyen la latencia de respuesta, la duración segmental, las pausas entre palabras (en frases), la probabilidad espectral y la frecuencia fundamental (F0). En la retroalimentación suprasegmental, en la que la entonación y la acentuación son importantísimas, también la F0 es una de las correlaciones acústicas principales. En la figura 23, se ofrece una reproducción de su diseño de reconocimiento del habla, donde se puede observar su complejidad.

Existen tres aspectos del habla que reciben atención especial: el desarrollo de vocabulario, la práctica conversacional y la pronunciación, incluyendo la práctica de pares mínimos y la anotación de la pronunciación.

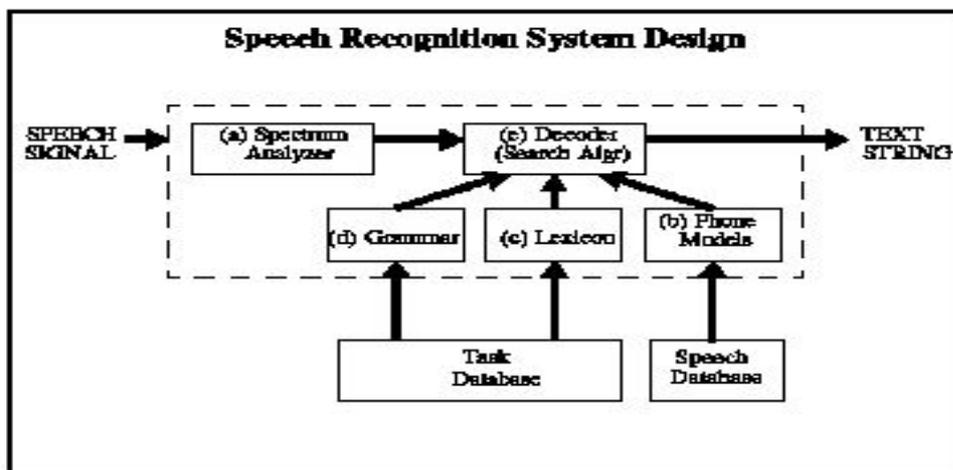


Figura 23. Ehsani y Knodt (1998:48). Diseño de un sistema de reconocimiento del habla

### **2.12.1 Los porcentajes de reconocimiento de voz**

Ainsworth y Greenberg (1996), en sus estudios sobre el reconocimiento automático del habla, dicen que trabajando de manera tradicional y con condiciones limitadas de vocabulario, suele dar un 98-99% de corrección. Ese porcentaje se reduce a un 88-95% cuando el material consiste en habla individual y teniendo siempre en cuenta que la grabación se efectúe en un ambiente acústico relativamente libre de ruido.

Sin embargo, bajo muchas condiciones que simulen la típica comunicación humana, la ASR funciona de manera mucho más pobre. El sistema informático de reconocimiento del habla suele fallar cuando se confronta con materiales hablados presentados bajo condiciones auditivas más realistas, en las cuales el retumbo y el alto nivel de interferencia acústica de fondo suelen ocurrir. Eso hace reducir el nivel de precisión en la palabra a un 20-50%.

Dos años más tarde, Greenberg *et al.* (1998), se refieren de una manera más técnica a la misma cuestión diciendo que cuatro canales espectrales distribuidos en un margen de habla de entre 0.3 a 6 khz son suficientes para que el ser humano pueda escuchar y decodificar materiales a nivel de frases con un 90% de precisión, aunque más de un 70% del espectro se pierde. El nivel de reconocimiento a nivel de palabras es relativamente alto (de un 60 a un 83%) cuando sólo se presentan dos o tres canales simultáneamente, a pesar de que la inteligibilidad de las mismas, presentadas de manera aislada, es de menos del 9%.

Bates (2003:1-2), en la introducción de su tesis sobre la variabilidad de pronunciación para modelos de reconocimiento de voz, en lo que se refiere a los datos de fiabilidad en reconocimiento de palabras con alta

calidad de sonido, comenta que el porcentaje alcanzado es mayor de un 90% en vocabularios de más de 50.000 palabras. Desafortunadamente, los resultados no son tan buenos cuando el habla es espontánea, donde se ha encontrado más de un 20% de errores al utilizar el teléfono y de más de un 35% con múltiples hablantes utilizando micrófonos ajustados a la cabeza.

### **2.12.2 Los reconocedores de voz de habla aislada y continua**

Faudez (2000:98) hace una distinción según la manera de enlazar las palabras en el habla:

- a) Reconocimiento del habla aislada: deben realizarse pausas suficientemente largas entre palabras (típicamente más de 200 ms), como si se realizara un dictado palabra por palabra. La ventaja es que resulta más sencillo detectar los inicios y finales de palabra que mediante habla continua, puesto que se han añadido silencios entre palabras. El inconveniente es que requiere la cooperación del locutor, que debe cambiar la manera normal de hablar.
- b) Reconocimiento del habla continua: es el sistema más complejo, ya que el reconocedor se verá afectado por efectos coarticulatorios. Un mismo fonema no siempre suena igual; depende de cuál haya sido el fonema anterior. La ventaja es que trabaja con la manera normal de hablar entre personas.

Según Ehsani y Knodt (1998), programas de reconocimiento de voz de manera continua como *Dragon's Naturally Speaking* o *IBM Via Voice*, por ejemplo, actúan con una precisión de reconocimiento de sólo un 60% a 80% (dependiendo del acento, ruido, tipo de producción, etc.).

### **2.12.3 Los reconocedores de voz con puntuación**

Llisterri (2003:14) trata los criterios de parecido o aceptación que se suelen aplicar en los programas educativos. En ellos, el reconocedor de voz ha de aceptar o rechazar enunciados con más o menos parecidos al original, y proporciona al alumno una puntuación, que no es otra cosa que la “distancia” entre el modelo que el usuario ha escuchado e intenta imitar y su propia producción. Dicha puntuación se establece numéricamente (con una escala de 1-10, de 1-5, etc.), pudiéndose en algunos casos (programas de *Auralog*), aumentar el criterio de evaluación en determinados programas para que el aprendiz intente realizar producciones mucho más precisas. Al referirse al proceso de reconocimiento en general y al específico, es decir, a los problemas localizados, he aquí una cita explicativa:

*“La comparación entre el enunciado almacenado y el original requiere también una deducción, por parte del sistema de reconocimiento, de los elementos que forman el enunciado, de modo que no únicamente se realiza una comparación global, sino que se lleva a cabo lo que se conoce como una “alineación”, es decir, una sincronización entre el texto correspondiente al enunciado y la señal sonora. Esto permite señalar al estudiante aquellos fragmentos de su producción en los que se produce una discrepancia más fuerte con respecto al modelo, de modo que puede concentrarse en ellos a la hora de repetir para mejorar la semejanza con el enunciado original”.*

Llisterri (2003:15)

Rypa y Price (1999) dicen al respecto que la anotación de la pronunciación en una actividad sólo se establece después de que se complete totalmente una lección y se hayan recogido suficientes datos con una puntuación aceptable. Para validar la puntuación automática de la pronunciación, en este sistema, se recoge tanto el habla nativa como la no nativa y se crea una base de datos de porcentajes realizada por expertos (por ejemplo, instructores lingüísticos).

### **2.12.4 La tecnología HMM (Hidden Markov Modelling)**

Varios autores (Ehsani y Knodt, 1998; LaRocca, 1999; Bernstein y Franco, 1996) han alabado la tecnología HMM (*Hidden Markov Modelling*) empleada en los reconocedores de voz. Dicha tecnología debe su nombre al matemático ruso Andrei A. Markov (1856-1922) y, según Rabiner (1989), tiene su origen en las publicaciones de una serie de documentos clásicos por Baum *et al.* (1966) y se implementó para aplicaciones de procesamientos del habla gracias a Baker (1975), en la Carnegie Mellon University (CMU) de Pittsburg, en Estados Unidos. Dicho modelo aplica estadísticas sofisticadas y computaciones probabilísticas al problema de enlazar modelos a nivel de fonemas y sílabas.

Para Ehsani y Knodt (1998), un reconocedor del habla HMM consta de cinco componentes básicos:

- a) Un analizador de señal acústica que computa una representación espectral de habla que se acaba de procesar.
- b) Un juego de modelos fónicos (HMMs) entrenados con grandes cantidades de datos de habla.
- c) Un léxico para convertir secuencias fónicas con fonemas y sílabas en palabras.
- d) Un modelo estadístico de lengua o enlace gramatical que defina la tarea de reconocimiento en términos de combinaciones de palabras legítimas a nivel de frase.
- e) Un decodificador, que es un algoritmo de búsqueda para computar el mejor enlace entre una producción hablada y su correspondiente palabra.

## 2.13 Los micrófonos

Por lo que se refiere a la calidad de la entrada de datos para una actuación óptima de reconocimiento, la señal de entrada del habla debe tener calidad acústica.

No sólo la interferencia de ruido puede afectar a la calidad del habla. Existen otros factores como el tipo de tarjeta de sonido, de micrófono empleado, o si el habla se efectúa mediante un pre amplificador. Algunas tarjetas de sonido tienen amplificadores internos, pero tienden a ampliar el sonido junto con el habla. La amplitud de la señal ha de ser ajustada cuidadosamente para su mejor reconocimiento.

Si observamos la figura 24, podemos ver los diversos mecanismos que influyen en un reconocimiento de la voz. Éste es un aspecto en el que el micrófono puede representar una tremenda diferencia en la actuación de reconocimiento. Muchos reconocedores actúan mejor cuando se emplea un micrófono ajustado a la cabeza con un cancelador de ruido. No sólo filtran el ruido exterior, sino que también, al estar colocados en la cabeza, se aseguran que la distancia entre la boca del hablante y el micrófono se mantiene constante y la amplitud es estable a través de las producciones.

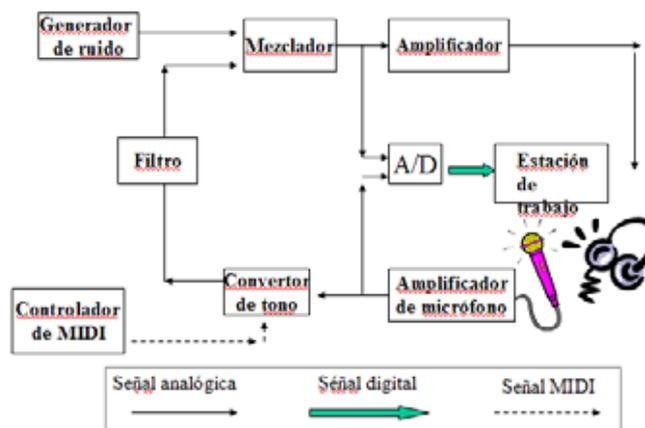


Figura 24. Kawahara (1996:15). Análisis de la sensibilidad de los parámetros del habla

Stibbard (1996) considera de crucial importancia la colocación del micrófono y dice que se ha de sostener el micrófono cerca de la boca, pero excesivamente, ya que algunos sonidos como los sibilantes o las vocales (altas – *loud*) producen una sobrecarga.

En este apartado se ha ofrecido una breve referencia teórica. Cuando tratemos las recomendaciones metodológicas, este aspecto se ampliará en el apartado 4.3.2

## **2.14 El análisis de errores**

Un sistema de reconocimiento de voz ha de ser capaz de reconocer los errores tanto segmentales como prosódicos. Al tratar la percepción del habla, Brett (1999) cita a Rost (1990), quien, al referirse a las características de la señal acústica, los fonemas y su articulación, comenta que la realización de fonemas ingleses está afectada por el contexto de los fonemas en los que aparecen, pudiéndose producir variaciones libres, asimilaciones, reducciones y elisiones. Las personas producen diferentes resultados acústicos dependiendo de su edad, sexo, región y características socioculturales, afectados también por la manera de expresarse (por ejemplo, placer, agitación, etc). No obstante, los fonemas son aún percibidos de manera exitosa. Igualmente, Cutler (1997), utilizando el *MARSEC Spoken Corpus*<sup>21</sup>, ha calculado que aproximadamente el 85% de las unidades polisilábicas en inglés contienen palabras incrustadas o encajadas, por ejemplo: ‘bracing’ contiene ‘brace’, ‘race’, ‘ace’, ‘racing’ y ‘sing’.

---

<sup>21</sup> MARSEC (Machine Readable Spoken English Corpus) está basado en muestras del habla inglesa en acento común del inglés británico del sur y ha sido desarrollado por Lancaster y IBM basándose en el SEC (Spoken English Corpus). La diferencia entre el primero y el segundo es que el MARSEC incorpora un segundo CD-Rom con grabaciones acústicas (el SEC carece de ellas) añadidas e incluye una alineación entre la palabra y el tiempo entre las transcripciones y la señal acústica.

Chanier *et al.* (1992:134) tratan el análisis de errores en los sistemas asistidos por ordenador. Este análisis, conocido en inglés como *parsing*, procesa el significado de los diversos comandos que utilizan algunas estrategias para reconocer una producción como errónea, localizar este error respecto a la gramática computacional y, a veces, poder remediarlo.

### **2.14.1 El alineamiento forzado de errores**

Un reconocedor de voz basado en un modo de alineamiento forzado (*forced alignment mode*) puede calcular los errores. El sistema enlaza el texto de la señal entrante como la señal usando información sobre el contenido lingüístico que ya ha sido almacenado en la memoria. Entonces, después de compararla con la nativa, se pueden identificar y localizar.

Varios autores tales como Horak (2001), Eskenazi y Hansma (1998) y Eskenazi (1999) se han referido a este tipo de tecnología. La siguiente cita resume su uso:

*"Given the sentence that the speaker was to say, the speech recognizer can return the scores of the words and the phones in the utterances. By comparing the speaker's recognition scores to the mean scores for native speakers for the same sentence pronounced in the same speaking style, errors can be identified and located. For prosody errors, only duration can be obtained from recognizer output. That is, when the recognizer returns the duration of the phones".*

Eskenazi y Hansma (1998:78)

### **2.14.2 Diversos estudios**

Sustarsic (2001) y Petek (1999) han elaborado estudios sobre áreas problemáticas en el reconocimiento de voz en inglés. Teniendo en cuenta algunos de los programas de reconocimiento de voz, tales como *Via Voice*

(patentado por IBM) o *Dragon's Naturally Speaking* (Dragon System Inc.), Sustarsic (2001) realizó diversas aproximaciones al análisis de los errores para trabajar pronunciación y encontró que la mayoría de errores de reconocimiento consistían en omisiones, adiciones y confusiones o malas interpretaciones. Este último punto se desarrolló en cuatro subapartados: vocabulario técnico, abreviaciones, vocabulario con nombres propios y vocabulario general.

Lo que es importante destacar es que, si el problema es el reconocimiento del habla de manera continuada -entiéndase aspectos suprasegmentales- la fiabilidad de los datos en los espectrogramas puede estar influenciada por los siguientes factores, que podemos resumir en la siguiente cita:

*“A word of caution though: speech recognition programs deal better with relatively slow, clear and “careful” (non-assimilated) style of speech, which might encourage students to train the “foreign language” word-by-word pronunciation – precisely the opposite of what we want them to achieve. On the other hand, this may not be such a serious problem if we instruct students to apply the more natural colloquial style already in the process of training the computer to their voices, following the simple philosophy that the computer is the tool of the user and not the other way round”.*

Sustarsic (2001:50)

Si bien el problema descrito es aquí mucho más preciso y no directamente relacionado con un programa de mejora de la pronunciación, debemos tener en cuenta que la producción por parte de los alumnos de frases de mayor o menor complejidad puede resultar difícil; debe realizarse un seguimiento del alumno y no permitir que se puedan producir fosilizaciones.

## PARTE II

### 2 MARCO TEÓRICO DEL MULTIMEDIA: La enseñanza asistida por ordenador y su relación con el análisis del habla

#### 2.1 Introducción al proceso audiovisual de comunicación

Al tratar con métodos multimedia es interesante y necesario revisar brevemente el proceso de comunicación audiovisual. Mallas (1977:25) - ver figura 8- comenta la necesidad que los educadores han de tener para conocer el circuito de comunicación cerrado, bajo cualquier fórmula operativa y funcional. Dicho circuito combina los elementos siguientes: emisor, mensaje, receptor, retroalimentación, contramensaje, receptor.

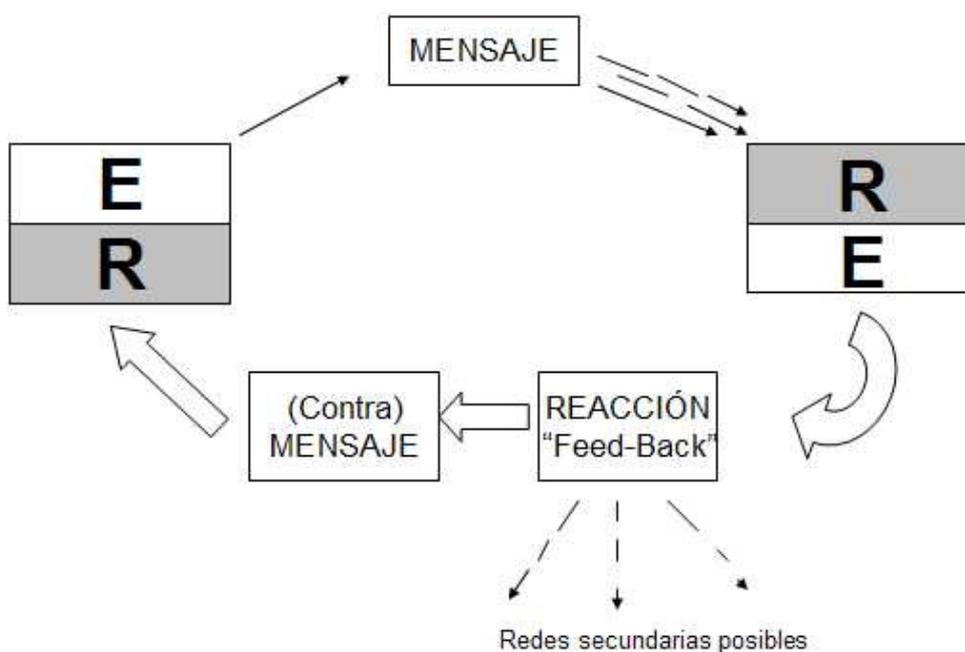


Figura 8. Mallas (1977:25). Circuito cerrado de comunicación

Todo circuito cerrado es bidireccional, pero también omni direccional, ya que en muchos casos se produce una formación de redes secundarias que generan una trama de interacciones que generalmente escapan a todo control. De este tipo de circuitos, Mallas resalta la “retroalimentación” o *feedback* que provoca la inversión de roles comunicativos; es decir, el cambio de flujo informativo y el receptor convertido en emisor. Finalmente, destaca que los audiovisuales que mejor se adaptan a la mayoría de funciones son los *subordinados al profesor* y conocidos también como “pequeños AV” (o medios audiovisuales). El ordenador entraría dentro de esta categoría.

También conviene comentar en qué se basa la irrupción tecnológica y sus necesidades de respuesta teórica, ya que un modelo de enseñanza basado en el uso del ordenador serviría para: enseñar mejor, más rápidamente, a mayor número de sujetos, con mejores instrumentos, paliando la escasez de docentes cualificados y facilitando mejor la tarea docente.

Es importante comentar y observar el llamado Cono de la Experiencia de Dale (1969), que se resumiría en la siguiente cita:

*“El mejor sistema de aprendizaje es, en principio, el de la experiencia. Se aprende y asimila mejor aquello que se hace que aquello que se ve; menos aquello que sólo se escucha, y prácticamente nada, aquello que sólo se oye”.*

Mallas (1977: 29)

En las siguientes páginas podemos observar el gráfico original de Dale (1969) así como un par de interpretaciones llevadas a cabo en varias universidades, que corresponden a las figuras 9, 10 y 11.

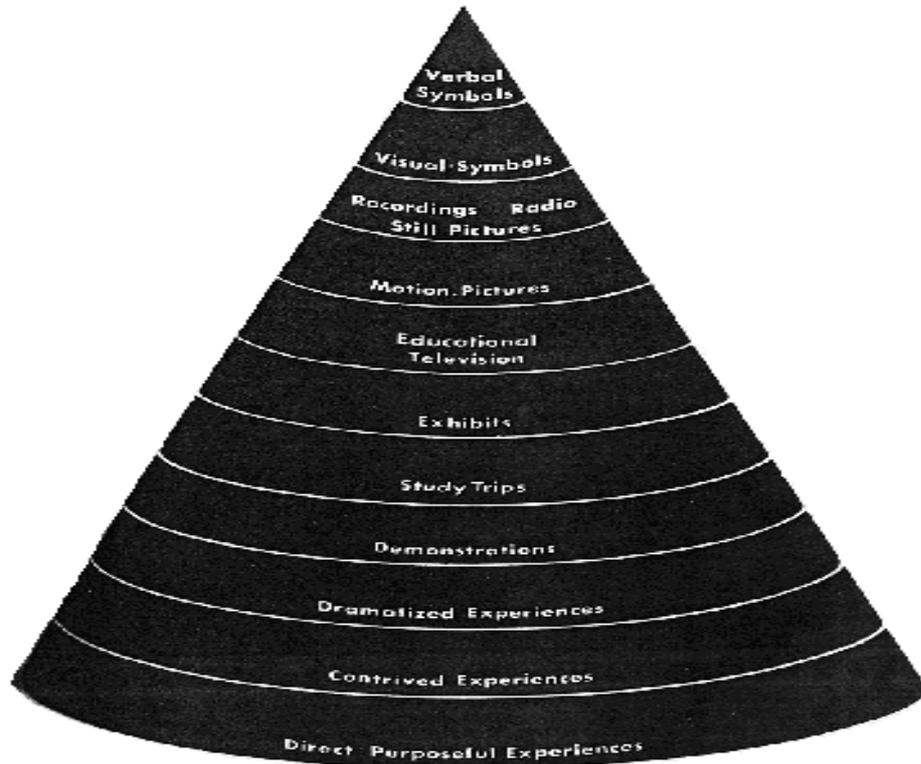


Figura 9. Dale (1969:107). Cono de la Experiencia  
<http://www.work-learning.com/chigraph.htm>

Table VI		
Dale's Cone of Experience		
<b>People generally remember:</b>	?	<b>Levels of Abstraction:</b>
<b>10% of what they read</b>	Read	<b>Verbal Receiving</b>
<b>20% of what they hear</b>	Hear words	
<b>30% of what they see</b>	Watch still picture	
	Watch moving picture	
<b>50% of what they hear and see</b>	Watch exhibit	<b>Visual Receiving</b>
	Watch demonstration	
<b>70% of what they say or write</b>	Do a site visit	<b>Hearing, Saying</b>
	Do a dramatic presentation	
<b>90% of what they say as they do a thing</b>	Simulate a real experience	<b>Seeing and Doing</b>
	Do the real thing	
	? ? ? ? ?	
<small>See Wiman &amp; Mierhenry, Educational Media, Charles Merrill, 1960, for reference to Edgar Dale's Cone of Experience.                      *Question marks refer to the unknown.</small>		

Figura 10. Chacón (2001). Interpretación del cono de Dale  
[http://www.willatworklearning.com/2006/05/people\\_remember.html](http://www.willatworklearning.com/2006/05/people_remember.html)

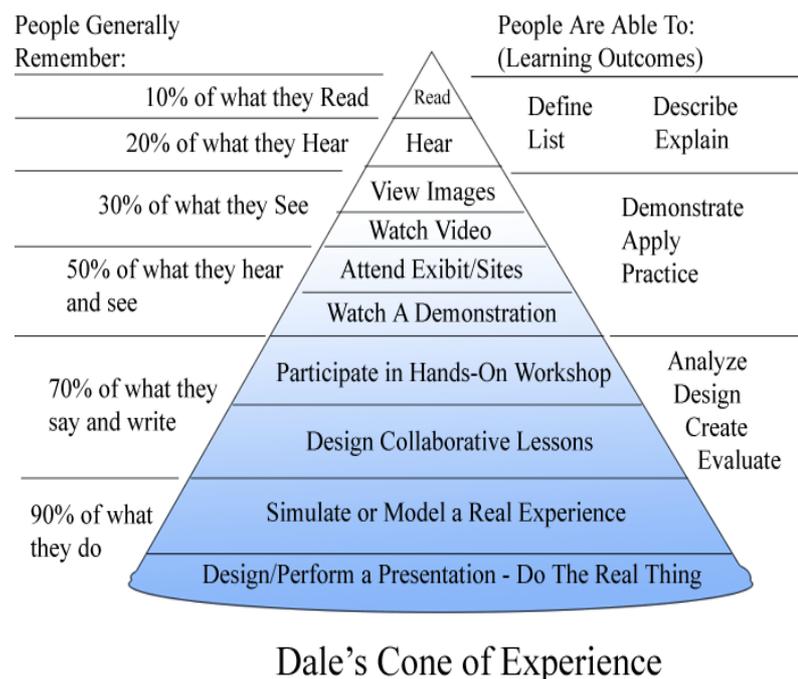


Figura 11. Pastore (2003). Interpretación del Cono de Dale  
<http://teacherworld.com/potdale.html>

La referencia de Chacón (2001), refiriéndose a Wiman y Meierhenry (1969), -ver figura 10- fusionando gráficos y situando números, plantea unos interrogantes que -como el pie de gráfico comenta (en inglés en la página anterior) se refieren al porcentaje inferior al 10% y superior al 90%, cuyo resultado es un punto desconocido.

Mallas (1977) ofrece, por otra parte, una interpretación ligeramente diferente, pero siempre en la línea de los dos gráficos anteriores. Según él, la relación directa con la realidad es mayor a medida que la información del triángulo se dirige hacia el extremo inferior (del 10% al 90%); de la misma manera aumenta la accesibilidad de la interpretación y, por lo que respecta al proceso contrario, aumenta el grado de abstracción y la versatilidad.

## 2.2 Definiciones del término multimedia

El concepto que aquí nos ocupa está relacionado con la irrupción tecnológica de una serie de medios de trabajo y, si bien tiene unos orígenes relativamente recientes, tiene -tal y como refiere Mallas (1977)- unos pre-orígenes *visuales didácticos* encaminados a apoyar la acción magistral (grabados, láminas, encerado, maquetas) que no son aún *medios tecnológicos*, ya que les falta el elemento transmisor: *la "maquina"*, sea un proyector de vistas fijas, etc.

La definición del término multimedia ha evolucionado mucho desde sus orígenes, como podremos observar en diversas citas a continuación. En los años 60 del siglo anterior, Ely (1963) se refería al término multimedia asociándolo con situaciones en las que se combinaban elementos de audio y vídeo de manera múltiple (incluyendo maquinaria considerada hoy en día obsoleta, tales como proyectores de diapositivas, reproductores de vídeo, monitores, cintas de casete, etc.) y otro tipo de *hardware* de una manera instructiva o no instructiva.

Evidentemente, los más de cuarenta años que han transcurrido desde los orígenes de las aplicaciones multimedia, han hecho que el concepto haya expandido su radio de acción y todo ello puede proporcionar al aprendiz una flexibilidad y un uso de estrategias cognitivas.

Alpiste equipara el término multimedia con el punto de encuentro de tecnologías y de "técnicas" que han venido madurando por su cuenta durante años para coincidir ahora en un lugar común y sitúa el efecto pre-multimedia y una referencia económica al coste, en su momento, diciendo:

*“A mediados de los 70 era un privilegio exclusivo de algunas instituciones disponer de algo que hoy nos parece tan elemental como un vídeo y un ordenador. El coste medio de esos equipos, si los comparamos con los aparatos domésticos de hoy, rondaba los dos millones de pesetas<sup>1</sup> (de 1975)”.*

Alpiste (1993:17)

De Bustos (1996), en su guía práctica para usuarios multimedia, se da cuenta de la dificultad de definir el término multimedia de una manera clara, señalando que no existe una definición aceptada por todos. Por tanto intenta hablar del concepto como idea y no como término a definir:

*“Multimedia es básicamente una forma de expresar información, o mejor dicho, son muchas formas de expresar información a la vez, ya que el ser humano es capaz de recibir información de muchas maneras a la vez, tantas como sentidos tiene (...) pero al mismo tiempo es capaz de dar información a cada sentido de muchas formas diferentes, por ejemplo, la vista puede recoger imágenes, colores, movimientos, etc., el sonido puede recibir sonidos armónicos, música, ruidos, etc. La forma de conjuntar todas esas formas de expresión, o las máximas posibles, e integrarlas en un todo, es lo que se conoce como multimedia”.*

de Bustos (1996:15)

Wise (2000) comenta la manera en la que las tecnologías multimedia han evolucionado desde las necesidades militares de producción de herramientas más precisas, y también comenta la problemática al definir el término:

*“Feldman [1997] has defined this sense of multimedia concisely as ‘the seamless integration of data, text, images and sound with a single digital information environment’. Such a definition concentrates on technologies and applications like electronic mail, on-line publishing, CD-ROM, the Internet, and digital television. However, we define multimedia in a less strictly technical way than this. Indeed, as far as the purposes of this book are concerned, such terms as ‘cyberspace’, ‘new media technologies’, ‘computer-based media’, ‘digital technologies’ and ‘information and communications technologies’ are adequate synonyms”.*

Wise (2000:2)

---

<sup>1</sup> Al cambio, 12.000 euros

El paralelismo del término multimedia con la tecnología de la información y la comunicación (TIC), o TAC (Tecnologías del Aprendizaje y del Conocimiento), un término de más reciente creación, así como de otro sinónimo, ‘Tecnología educativa’, serán comentados más adelante. Una definición más precisa, actualizada y desmenuzada en diversos apartados, la encontramos a continuación:

*“Multimedia is the combination of multiple forms of media in the communication of information. Multimedia enables people to communicate using integrated media: audio, vídeo, text, graphic, fax and telephony. The benefit is more powerful communication. The combination of several media often provides richer, more effective communication of information or ideas than a single media such as traditional text-based communication can accomplish. Multimedia communication formats vary, but they usually include voice communication (vocoding, speech recognition, speaker verification and text-to-speech), audio processing (music synthesis, CD-ROMs), data communications, image processing and telecommunications (...) and networks. Multimedia technology will ultimately take the disparate technologies of the computer, the phone, the fax machine, the CD player, and the vídeo camera and combine them into one powerful communication center”.*

Newton (2004: 544)

El gran potencial actual de la tecnología multimedia es la tecnología digital que, tal como podemos ver en el siguiente gráfico de Goodman (2004), figura 12, resume muchos de los aspectos ya tratados:

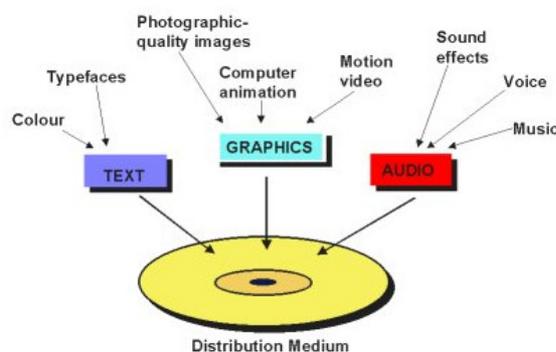


Figura 12. Goodman (2004) Tecnología Multimedia  
<http://www.deakin.edu.au/~agoodman/sci204/lecture1.html>

Alpiste (1993:25) sitúa el primer ordenador personal merecedor del calificativo multimedia al Macintosh II, en su día con sonido y gráficos en 256 colores y con un interfaz amigable. Si bien otros autores, como Wise (2000:44-45), precisan más modelos de maquinarias importantes (Altair –el primer ordenador personal en venderse- y empresas como Apple e IBM PC).

## **2.3 Términos relacionados con Multimedia**

El término multimedia se ha asociado con diversos vocablos y diversos tecnicismos que conviene precisar. A continuación, desarrollaremos términos que merecen su atención, ya que tienen relación con esta palabra, como son: el tutor inteligente, la inteligencia artificial, la tecnología educativa, la interactividad, hipermedia y multimedia interactiva.

### **2.3.1 El tutor inteligente y la inteligencia artificial**

El concepto de *entorno con un tutor inteligente* para referirse al mismo tema podemos definirlo así:

*“An intelligent tutoring environment consists of multiple components including presentation of content material and student interactivity opportunities. The point is that an epistemological framework, instructional theory, and a set of related pedagogical strategies should be at the heart of technology-based instructional design products”.*

Miller (2003: 2)

El hecho de que el constructivismo esté detrás de este *tutor inteligente* hace que se incluyan diversas estrategias formativas, en las que se combinan diferentes aspectos, tales como una aproximación a la

solución de problemas, una inclusión basada en el aprendizaje de un caso, el uso de contextos auténticos, modelos a seguir y un entrenamiento en dichos aspectos, todos ellos conjuntados con el uso de los medios multimedia.

Pennington (1989) utiliza el término *Artificial Intelligence* (AI) para hablar del *software* que hace que el ordenador se comporte de una manera inteligente, comprenda, aprenda y responda. La información restringida a través de la cual opera el ordenador es también conocida como *microworld*. Dicha autora también sugiere los principios de motivación intrínseca, interactividad y eclecticismo y afirma que la cuestión de la comunicación con un ordenador es una de las más intrigantes en CALL.

El llamado Turing Test (ver Saygin *et al.* 2001), basado en el modelo utilizado por Turing en los años 1950 y su pregunta básica acerca de “¿Pueden pensar las máquinas?”, puso en estudio a una persona en conversación con dos mecanismos, uno de los cuales era conducido por un ordenador y otro por un humano. Si no se podía distinguir cuál era operado por el ordenador, se decía que el ordenador había pasado dicho test.

### **2.3.2 La tecnología educativa**

La siguiente cita nos sitúa en un contexto histórico acerca de este tema:

*“La función histórica de la tecnología educativa es más un proceso que un producto. No importa qué nivel de sofisticación puedan tener los métodos de la instrucción, se debe hacer siempre una distinción precisa entre el proceso de desarrollar una tecnología educativa y el uso de ciertos productos o medios dentro de una tecnología de instrucción particular”.*

Saettler (1990: 4)

En relación con la definición de este término, algunos consideran la tecnología educativa como una rama de la teoría y de la práctica educativa que está afectada “primariamente con el diseño y uso de mensajes que controlan el proceso de aprendizaje”.

Ely (1963) hace una clara distinción entre mecanismos y equipos o entre la *ciencia física* de la tecnología educativa y la *ciencia conductista* de la tecnología educativa, y considera *el proceso* como el saber y poder realizar una destreza, y considera el *producto* como el equipo y los materiales utilizados en el proceso. Otra definición de Tecnología educativa sería:

*“Tecnología educativa es un proceso complejo integrado que implica a la gente, los procesos, las ideas, los mecanismos y las organizaciones, para analizar problemas, concebir, implementar, evaluar y gestionar soluciones a esos problemas que implican todos los aspectos del aprendizaje humano”.*

Saettler (1990:6)

Conductistas como Skinner, en la década de los años 1950, concluyeron que eventos mentales tales como pensamientos, imágenes o conocimiento no tenían lugar en la ciencia de la psicología porque no se podían observar directamente. Las posiciones conductistas en psicología empezaron a declinar a finales de los años 50 del siglo pasado, irónicamente a medida que el impacto de la tecnología educativa se aceleraba. Las nociones de Skinner de refuerzo y su aplicación a las máquinas de enseñanza e instrucción programada habían empezado a influenciar el desarrollo de una tecnología de instrucción basada en una ciencia. La instrucción asistida por ordenador (en inglés CAI – *Computer-assisted instruction*) fue influenciada por esta tendencia. La orientación conductista llevaba a un currículum que era programado paso por paso en unidades pequeñas y que hacía hincapié en productos educativos observables.

Por otra parte, el concepto cognitivo de la tecnología educativa, intenta entender los procesos internos de conducta y poner el énfasis en el conocimiento más que en la respuesta. Se ve al aprendiz no como ser pasivo, sino activo, constructivo y participante. Este modelo empezó a reemplazar al conductista a principios de los años 80 del siglo pasado.

Es interesante citar a Richards *et al.* (1998) quienes se refieren a este tipo de estilos definiéndolos como comportamientos psicológicos que sirven como indicadores relativamente estables de cómo los alumnos perciben, se relacionan y reaccionan ante el entorno de aprendizaje; dichos estilos están relacionados íntimamente con tipos distintos de personalidad.

Willing (1988:101) realizó un estudio a gran escala con alumnos de inglés en Australia e investigó cómo las diferencias en estilos cognitivos afectaban las preferencias de los alumnos en seis áreas distintas:

1. Preferencia por ciertas actividades de clase
2. Preferencia por ciertos tipos de comportamiento del profesor
3. Preferencia por ciertos agrupamientos
4. Preferencia por destacar ciertos aspectos de la lengua
5. Preferencia por modos sensoriales concretos, como el aprendizaje visual, auditivo y táctil
6. Preferencia por formas concretas de aprendizaje personal, fuera del aula

Fue precisamente en estos cuestionarios sobre preferencias de aprendizaje donde la preferencia más citada, con un 62%, fue el gusto por practicar los sonidos y la pronunciación.

### **2.3.3 La interactividad**

Otros autores introducen el concepto de interactividad al hablar de multimedia, en la cual el usuario ejerce el control sobre las decisiones a tomar:

*"The most commonly accepted definition of multimedia supports the concept of computer-driven interactivity and a learner's ability to determine and control selection and sequence of content".*

Burton *et al.* (1995: 346)

Matchell y Elliot introducen términos paralelos y relacionados entre sí para hablar de *multimedia* y de *hypermedia*:

*"Interactive multimedia has characteristics of both hypertext and hypermedia. Interactive multimedia should incorporate the following elements: motion video and/or animation; voice, data, graphics, still images, audio; sound and text. Successful incorporation of one or all of these elements can be determined by the degree to which these elements augment, support, and reinforce the achievement of learning objectives and the level of interactivity. However, inappropriate or illogical integration of one or more of the aforementioned media elements may serve to confuse the learner and reduce the possibility of achieving terminal instructional objectives".*

Matchell y Elliot (1991:11)

Más adelante trataremos la interactividad con más detalle.

### **2.3.4 Hipermedia**

La fusión de este término puede ser definida así:

*“The term hypermedia is the merging of multimedia and hypertext. There exists considerable discussion about how both hypertext and hypermedia promote their design as replicating brain functioning in reference to how people learn, access, and use information and ideas. Research has associated relevant psychological phenomena such as cognitive learning principles with the instructional efficacy of hypermedia”.*

Borsook y Higginbotham-Wheat (1992)

Y por lo que respecta a la definición de hipermedia, en la siguiente cita se nombran tres definiciones diferentes, siendo la tercera la más relacionada con los aspectos educativos aquí tratados:

*“A way of delivering information that provides multiple connected pathways through a body of information. Hypermedia allows the user to jump easily from one topic to related or supplementary material found in various forms, such as text, graphics, audio or video”.*

*“Another definition found is: Non-linear media, of which multimedia can be a form. Just as hypertext is a non-sequential, random-access arrangement of text, hypermedia is a non-sequential, random-access arrangement of multiple media such as video, sound and computer data”.*

*“A third definition: Hypermedia is a type of authoring and playback software through which you can access multiple layers of multimedia information related to a specific topic. The information can be in the form of text, graphics, images, audio or video”.*

Newton (2004: 404)

## **2.4 Elementos de un sistema multimedia**

En este apartado vamos a explicar los diferentes componentes que son necesarios en cualquier sistema multimedia y que conviene conocer en lo que respecta a su funcionamiento.

## 2.4.1 Componentes básicos

Alpiste (1993) integra los siguientes componentes básicos de un sistema multimedia:

- Procesamiento multimedia: un microprocesador con cierta capacidad para gestionar el control de imágenes y sonidos.
- Memorias ópticas: todo tipo de procesamientos digitales, entre los que llega a contabilizar más de 20 distintos.
- Interfaz hombre máquina: dispositivos de interacción con los programas. El interfaz sería un vínculo mecánico o eléctrico conectando dos o más piezas de equipos. Existen numerosos dispositivos para la interacción con los programas. Los más difundidos son los teclados y los distintos sistemas de punteo (ratón, pantalla táctil, mandos para consolas o *joystick*, etc). Los lectores de códigos de barras, reconocedores de caracteres y reconocedores de voz son también interfaces.
- Sistemas de salida: donde se presentan los programas.
- Sistemas de entrada: cualquier sistema de captación de imágenes, sonidos o datos que un sistema multimedia pueda aprovechar.
- *Software* de desarrollo: herramientas de “autoría”, programas de dibujo, animación, creación o gestión de recursos multimedia. En el siguiente subapartado podemos ver una ampliación sobre el origen y desarrollo de estos programas, que permiten una flexibilidad adicional en las áreas de:

- Visualización en el monitor: permiten gráficos más atractivos e imágenes.
- Juicio de respuesta: aceptan más de una respuesta.
- Registro de datos (*recordkeeping*): almacenan los diferentes tipos de actuaciones de los alumnos.

- Ramificación de datos (*Branching or looping*): pueden variar los contenidos para usuarios distintos. Algunos programas permiten un convertidor de señal. La tecnología ha de poder producir esa variación de contenidos para también permitir encontrar una actividad y poder volver al punto inicial.

Podemos distinguir cuatro tipos de aplicaciones de aprendizaje con autonomía:

1. Estructurada por el profesor / contenido fijo: normalmente el típico *drill software*, o batería de ejercicios.
2. Estructurada por el aprendiz / contenido fijo: el aprendiz escoge su propio camino. Los alumnos pueden trabajar ejercicios de pronunciación, ver y oír palabras y frases de la lección, grabar su propia pronunciación, escuchar el modelo las veces que quieran y re-grabar para mejorar.
3. Estructurada por el profesor / contenido variable<sup>2</sup>.
4. Estructurada por el aprendiz / contenido variable (ver pie de página 2): conviene también destacar la enorme inversión de horas en este tipo de aplicaciones<sup>3</sup>.

---

<sup>2</sup> Los contenidos variables vendrían a suponer variaciones de contenidos y mayor flexibilidad de trabajo. Si bien en pronunciación, la mayoría de programas estándar suelen formar parte de los dos primeros modelos comentados anteriormente.

<sup>3</sup> Brown (1991) hace una estimación de que por cada hora de material de aprendizaje multimedia se han de invertir entre 50 y 217 horas de trabajo personal de creación de materiales interactivos, si bien no hace ninguna precisión más sobre este dato.

## 2.4.2 El Software de autoría

Antes de cerrar este apartado, debemos hacer una mención a un tipo de programa educativo conocido como “de autoría” (en inglés, *authoring software*) en el cual el contenido es modificable y adaptable a las necesidades del usuario y regulado por el profesor. Aunque en el aspecto que aquí tratamos, como es la enseñanza de la pronunciación, no hay prácticamente ningún programa que se pueda adaptar (a excepción de *Pro-nunciation*), es necesario comentar sus orígenes.

El llamado *Authoring Language* es un lenguaje diseñado específicamente para producir *software* educativo. Según Wachman (1999), éste consiste en programas que el usuario puede adaptar añadiendo datos<sup>4</sup> que encajen con sus necesidades específicas en función del grado de dificultad, interés, cultura, etc., permitiendo un alto grado de autonomía. Davies *et al.* (2005) engloban los mismos usos, añadiendo recursos como imagen o grabaciones de audio y vídeo, que se incorporan como materiales de aprendizaje por medio de plantillas.

Según Gros (1994), el *software* de autoría, aparece a finales de la década de los ochenta del siglo pasado dirigido a diseñadores instructivos, para que no se tenga que utilizar ningún lenguaje de programación. Para ello se han de automatizar algunas de las decisiones que hasta el momento eran realizadas por el diseñador del programa. El diseño instructivo es una tarea compleja y, por consiguiente, su automatización no es un problema trivial.

Gros (1994) distingue tres tipos de automatización: el primero sería la automatización del diseño *y/o producción del software educativo*, la cual responde a la necesidad de cubrir una importante demanda social: el desarrollo de programas de formación de adultos (a distancia, formación

---

<sup>4</sup> Según él, normalmente en forma de texto, aunque por supuesto, con la mejora de los sistemas informáticos es posible incluir audio y vídeo.

en la empresa, etc.). La mayor parte de los programas de formación utilizan tecnología informática sin necesidad de tener que conocer un lenguaje de programación específico. Se trata de automatizar algunas de las decisiones relativas al diseño, la producción del programa, las estrategias de enseñanza, etc. Algunos programas se centran únicamente en la automatización del diseño sin entrar en la producción del programa. Otros, incluyen no sólo el diseño sino también la producción.

Los otros dos sistemas de automatización son: la automatización de *sistemas de transmisión del conocimiento* (tutores inteligentes) y la automatización de *sistemas de ayuda al aprendizaje* (denominados herramientas cognitivas que ayudan a la adquisición de estrategias de aprendizaje).

Estos primeros programas se fundamentaron en el método de producción basado en las teorías del aprendizaje de Gagné (1987), denominado *sistema de desarrollo instructivo*. Este modelo parte de la idea de que, antes de emprender un diseño instructivo, las metas instructivas y los resultados de aprendizaje deben quedar perfectamente definidos en base a cambios observables. El proceso de diseño se divide en cinco etapas: análisis, diseño, producción, implementación y mantenimiento.

En la fase de análisis, el diseñador ha de estudiar los requisitos para la instrucción: contenido, características de los alumnos, niveles de ejecución, etc. En la fase de diseño, hay que desarrollar las especificaciones instruccionales y el test de enseñanza. Mientras que estas fases están dominadas por conceptos abstractos, las dos fases siguientes, producción e implementación, requieren acciones más prácticas. Es necesario crear los materiales de enseñanza y utilizarlos. La fase de mantenimiento permite la evaluación y puesta al día del producto. Es en este aspecto donde la inclusión de palabras, imágenes o de cualquier aspecto modificable entra en acción.

## **2.5 Principios generales que deben regir una aplicación multimedia**

Bou (1997) enumera una serie de principios que deben de regir cualquier aplicación multimedia. Estos son el principio de múltiple entrada, interactividad, libertad guiada, retroalimentación, vitalidad, necesidad, atención y unicidad. Todos ellos son ampliados a continuación en este apartado.

### **2.5.1 El principio de múltiple entrada**

Bou (1997) habla de este tipo de aplicaciones refiriéndose a diseños con perfil de destino, para ser utilizadas por un determinado tipo de usuario. Así pues, distinguimos tres parámetros distintos para almacenar una información en nuestra memoria:

- a) El factor cognitivo: si su complejidad se puede asumir por nuestras destrezas cognitivas.
- b) El factor afectivo: el impacto de esa información en nosotros. Además, teniendo en cuenta las teorías de Krashen<sup>5</sup>, también existe otro principio según el cuál los aprendices tienen un *filtro afectivo*, que es una barrera imaginaria que, de manera involuntaria, les impide aprender cuando están estresados, faltos de motivación o inseguros.
- c) El factor de la experiencia previa: cómo reaccionamos ante una información similar y en qué esquemas mentales o ideas vamos a integrarla.

---

<sup>5</sup> Brett (1999), en relación a este aspecto, es partidario de dejar explorar libremente a los alumnos aquellas áreas que puede que no les guste admitir que no han entendido.

Por consiguiente, según el principio de múltiple entrada, para lograr una buena comunicación hay que utilizar todos los canales posibles.

### **2.5.2 El principio de interactividad**

Se trata de un recurso muy importante propio de los sistemas informáticos. En consecuencia, el diseño de interacción<sup>6</sup> en una aplicación multimedia debe regirse por unas reglas genéricas que se tratan a continuación:

- a) El refuerzo del mensaje. Podemos incluir quizás en este aspecto lo que Brett (1998b) llama la repetición múltiple, es decir, la posibilidad de repetir el mensaje cuantas veces se desee.
- b) El no utilizar la interactividad supone un desperdicio de potencialidad y de intervención del usuario.
- c) Participación activa, no repetición de gesto. Alpiste (1993) precisa este término aludiendo al ejemplo práctico de utilización de la tecla INTRO para pasar pantallas y, evidentemente, no lo considera como interacción, sino repetición de gesto.
- d) No limitar el mensaje al esquema usuario – máquina: puede promover que las personas dialoguen y cooperen y no debe fomentar un aislamiento personal.

---

<sup>6</sup> Barney y McCabe (2001:4) citan a Vygostky (1962) en relación a la interacción, quien subraya la necesidad de la interacción social para aprender, y mantener esa interacción es clave en el desarrollo del lenguaje del niño. La interacción requiere negociación del aprendizaje y nosotros determinamos lo que nuestros interlocutores saben y necesitan saber para que la comunicación sea efectiva. La interacción no es sólo saber la respuesta correcta o incorrecta, y la interacción real sólo puede existir entre dos seres humanos; por ello el papel del profesor es crucial.

- e) Posibilidad de obtención de un registro de datos descriptivos de conducta del usuario.

### 2.5.3 El principio de libertad (guiada)

El objetivo del guionista es ocultar la estructura interna del programa educativo y que el usuario piense que navega libremente mientras está inmerso en un esquema de etapas predeterminado.

Brett (1998b) habla del concepto de *autonomía* o *del control por parte del alumno*, si bien el principio al que estos dos términos pertenecerían sería el mismo. Healey (1999b) detalla las condiciones que mejoran la autonomía y el aprendizaje en relación al aprendiz (ver figura 13):

- Grado de auto-motivación
- Preferencia por un estilo independiente
- Conocimiento de cómo se trabaja mejor
- Conocimiento de cómo uno necesita aprender

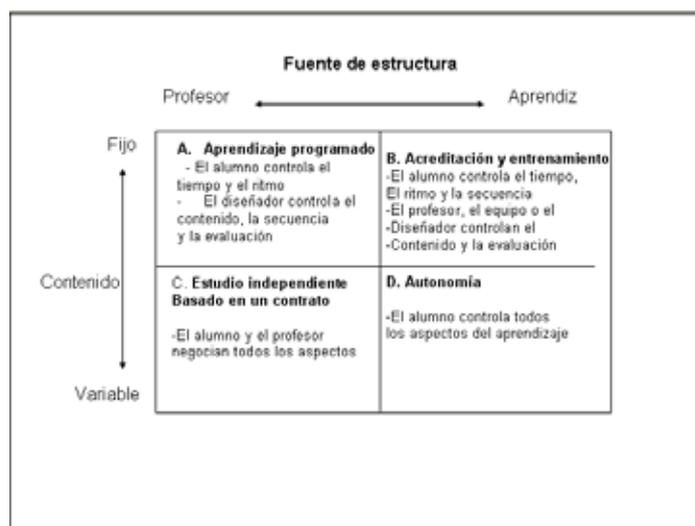


Figura 13. Healey (1999b:393). Parámetros para un aprendizaje autónomo

La figura anterior nos ayuda a ver los diferentes aspectos que se pueden dar dentro de los grados de autonomía entre el profesor y un aprendiz.

#### **2.5.4 Principio de retroalimentación**

Se denomina así a la idea de un sistema que genera información y que se utiliza para corregir su funcionamiento. Para ello hay que tener en cuenta qué tipo de información se recoge, cómo se presenta, a quién se dirige y cómo se procesa. Conviene resaltar más esta idea refiriéndonos a la teoría conductista de adquisición de lenguas de Skinner (1957), según la cual se suponía que el aprendizaje se producía a través de un proceso de retroalimentación o *feedback* de estímulo y respuesta, es decir, imitando el *input* expuesto y reforzándolo. Según esta teoría, la audición tenía un papel relevante en el aprendizaje.

Bou (1997) plantea una idea bastante similar situando lo que él llama la *información elaborada* (ver figura 14 en la página siguiente) como paso siguiente al módulo de análisis de respuestas.

Brett (1999), en relación a la retroalimentación del CALL, dice que este factor está siempre presente, es instantáneo, se aplica a cada respuesta incorrecta, a menudo proporciona la oportunidad al hablante de autocorrección y es privado. En las conclusiones de su tesis doctoral sobre las aplicaciones multimedia y su relación con la comprensión auditiva, Brett concluye diciendo que la retroalimentación de comprensión de una lengua en tiempo real es beneficiosa, así como también el que los aprendices se den cuenta de sus propios errores.

De todas maneras hay que seguir investigando sobre qué tipo de retroalimentación es la más beneficiosa.

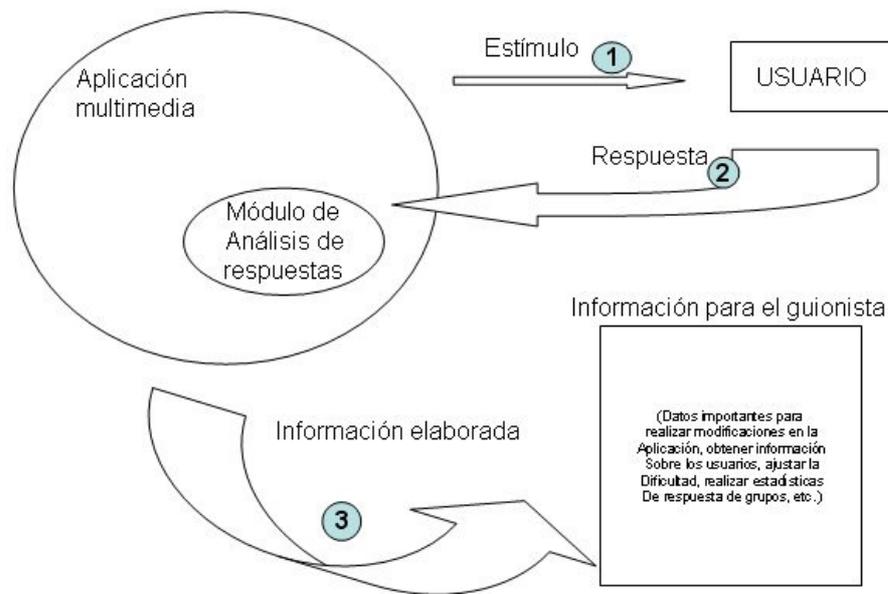


Figura 14. Bou (1997:40). La retralimentación: módulo de análisis de respuestas

### 2.5.5 Principio de vitalidad, necesidad y unicidad

Bou (1997) comenta que toda pantalla está viva. Es decir, el usuario tiene que percibir la aplicación como un mundo al que se asoma. Siempre sucede algo, aunque el usuario no haga nada. Tres consideraciones importantes acerca de este principio servirán para aclararlo un poco más:

- Los iconos animados resultan agradables.
- También lo son los iconos que responden instantáneamente.
- Son desagradables aquellos botones que no van a responder.

A su vez, el principio de necesidad se basa en el hecho de que la aplicación tenga un sentido, que sirva para algo y que sea multimedia. La aplicación viene a resolver un problema cuya solución percibimos inmediatamente que requiere un diseño multimedia. Toda producción

que no nazca de estas dos condiciones es gratuita. Es aquí donde la comodidad y la accesibilidad tienen protagonismo.

Según el principio de unicidad, la misión principal de la persona que coordina una aplicación multimedia es hacer que el producto final parezca obra de una sola persona, que tenga una unidad, una coherencia y un diseño de similares características. Bou (1997) comenta este punto en el sentido de la profesionalidad. Teniendo en cuenta el tipo de aplicación podemos prescindir o no de este principio, pero ha de exigirse una técnica y un estilo.

## **2.5.6 Principio de atención**

El objetivo de las aplicaciones es mantener lo que se llama *la atención sostenida*, es decir, conseguir que el receptor mantenga una actitud continua de expectación ante la aplicación. Para ello se dispone de dos factores descritos en los siguientes puntos.

### **2.5.6.1 La atención cognitiva**

Hemos de tener en cuenta la relevancia de la información y su organización. Ellis (1990:184) resume la perspectiva cognitiva de cómo los alumnos ganan control sobre el conocimiento de otra lengua gracias a la práctica libre que proporciona que el conocimiento de procesamiento controlado se convierta en automático.

Chapelle (1998, 2001) comenta las condiciones que pueden influir en la asignación de la atención durante la realización de tareas en una segunda lengua:

- 1) Interacción modificada: que, según Larsen-Freeman y Long (1991:144), juega un papel importante en la negociación del significado y ayuda a que la entrada de datos sea comprensible mientras aún contiene elementos lingüísticos desconocidos. Han de trasladar al aprendiz a un objetivo y deben centrarse en el lenguaje. La siguiente cita nos resume una buena explicación:

*"These interactions [with the computer] need to move the learner toward a task goal and stop progress along the way to focus on the language. Interaction can be accomplished through mouse clicks and hypertext links or through the learner's linguistic output and the software's speech recognition system. In the latter case, of course, the quality of the interaction will depend in part on the speech recognition capability of the software".*

Chapelle (1998:28)

- 2) Producción modificada: la identificación y corrección de errores dirige la atención a las formas lingüísticas y a la precisión.
- 3) Presión de tiempo: cuando no hay presión de tiempo es probable que exista más atención.
- 4) Modalidad: la comunicación escrita permite más oportunidades para que se forme la atención, mientras que el lenguaje oral suele ocurrir bajo la presión temporal, para que se consiga fluidez.
- 5) Ayuda: cuando los aprendices tienen ayuda sobre algunos aspectos del lenguaje, es probable que tengan más recursos relativos a más aspectos.
- 6) Sorpresa: un elemento sorpresa podría hacer que la atención decreciera, ya que interrumpe los planes y hace que se de importancia a esa sorpresa, pero esta hipótesis dependería de la naturaleza de esa sorpresa.

- 7) Control: el control de varios aspectos de esa tarea por el profesor o el aprendiz puede hacer que se incite a dar más enfoque a la forma, pero se necesita más investigación sobre qué tipo de cuestiones hay que controlar.

### **2.5.6.2 La atención afectiva**

La atención afectiva está basada en el lazo afectivo del usuario con la aplicación. Robinson (1995) define tres tipos de atención:

- Los procesos que están relacionados con escoger la información que se ha de procesar y mantener en la memoria.
- La habilidad de procesar la información.
- El esfuerzo mental en hacerlo.

Schmidt (1990) ve cuatro aspectos de conocimiento relacionado con el aprendizaje de lenguas: intención, atención, conciencia y control. Dentro del campo de conciencia, hace referencia al término “*darse cuenta*” y comenta que se aplica a todos los aspectos del lenguaje, incluida la pronunciación. Aunque esto no quiere decir que automáticamente un ítem del lenguaje vaya a ser recordado y aprendido, sí que este *darse cuenta* es una condición necesaria y suficiente para que el *input*<sup>7</sup> (la entrada de datos) pase a ser *intake* (datos internalizados y adquiridos).

Johnston (1999: 336) se refiere a las responsabilidades del profesor citando a Brown (1994: 93), un autor que ha dado prioridad a temas afectivos, entre los que se incluyen:

---

<sup>7</sup> Este concepto se ampliará en el apartado 2.7.2

- a) Se ha de reconocer la fragilidad del *ego del lenguaje* en una segunda lengua.
- b) La auto-confianza del aprendiz es un elemento crucial en su camino al éxito.
- c) Para triunfar hay que asumir riesgos.
- d) Los encuentros del aprendiz de lenguas con la cultura de la lengua de estudio son una parte vital del proceso de aprendizaje de lenguas.

### **2.5.6.3 La tecnofobia y la ansiedad**

Otro concepto tratado por Johnston (1999) en relación al ambiente sería el concepto de tecnofobia (falta de confianza o desagrado hacia la tecnología, considerada otra forma de ansiedad). Se han tratado también otros conceptos similares, como la *computerphobia*, o ansiedad hacia el ordenador, también llamada *techno-stress*.

Weil y Rosen (1997) se refieren a una serie de estudios y encuestas en las que del 30-80% de la gente muestra algún tipo de ansiedad utilizando ordenadores. Una profundización en el concepto de tecnofobia nos muestra sus características:

- Confusión causada por términos informáticos
- Sentimiento de quedarse atrasado cuando el resto utiliza el ordenador
- Miedo a que trabajar con el ordenador sea demasiado difícil
- Miedo a cometer errores que no se puedan luego corregir
- Una ansiedad más generalizada del papel de los ordenadores en la vida actual
- Problemas diarios en el uso del ordenador: pérdida de datos
- El efecto de deshumanización de los ordenadores
- Sentimiento de falta de control

Existen estudios de Beauvois (1993, 1994) en donde, gracias al uso de la CMC (*Computer-Mediated Communication*) en LAN (*local area network*) en laboratorios y clases de LE, se reducía la ansiedad de la actuación en la clase y se detectó una mayor producción en una clase típica LAN que de la manera oral en clases sin ordenadores. Dichos estudios se llevaron a cabo con un paquete de *software* denominado InterChange.

Young (1999:31) cita a Price (1991) refiriéndose al problema de la pronunciación relacionándolo con la ansiedad en clase. Si bien una pequeña intimidación es necesaria, esta autora en diferentes entrevistas notó que los alumnos parecían muy preocupados acerca de cometer errores en su pronunciación en LE y también deseaban en particular desarrollar un acento que se aproximara al de un hablante nativo. La reticencia a hablar en público era debida principalmente al temor a cometer errores.

### **2.5.7 Estudios con elementos multimedia y aplicaciones con ordenadores**

Según Mayer (1997:14), el aprendizaje multimedia se produce cuando los alumnos reciben información presentada en más de un modo. Al definir el aprendizaje multimedia, es importante distinguir entre *delivery media* (el sistema utilizado para presentar la instrucción), *presentation modes* (formato usado para representar la presente instrucción) y *sensory modalities* (los canales de procesamiento de la información que un aprendiz emplea para procesar la información, tal como el acústico y/o el visual). Mayer predice un efecto positivo en aquéllos que coordinan una explicación visual con una verbal. Los efectos positivos son más importantes en alumnos con bajo conocimiento previo que en aquellos con un alto conocimiento del campo en cuestión. La razón<sup>8</sup> que ofrece es

---

<sup>8</sup> Cabe comentar que en esta investigación, el campo de estudio era el de la mecánica, y se evaluaron actitudes de alumnos que recibieron instrucción multimedia mediante vídeo y audio. El campo que aquí nos ocupa tiene connotaciones diferentes.

que los alumnos de nivel alto es más probable que puedan crear sus propias imágenes mentales cuando se presenta una explicación verbal y que puedan construir conexiones entre las representaciones verbales y visuales.

Brown y Howlett (1994) describen el trabajo de desarrollo y evaluación llevado a cabo en el NCET (National Council for Educational Technology), en Gran Bretaña, que permitió obtener una serie de evidencias. También existían una serie de contradicciones; pero una cuestión quedó clara: no tenía ningún sentido evaluar las contribuciones que puede hacer la TI (tecnología de la información) a menos que los niños y niñas pudieran trabajar con ordenadores de manera regular. A continuación podemos resumir una serie de cuestiones que sirven para ver el valor de la TI en el entorno educativo:

- 1) La TI puede proporcionar un entorno de aprendizaje seguro: el miedo a equivocarse obstaculiza los primeros pasos, pero experimentar con nuevas ideas y aprender de los errores es importante. El ordenador proporciona un entorno seguro. Permite detenerse a pensar sobre un problema sin molestar al resto de la clase. El ordenador no se cansa y no es impaciente ni ridiculiza los esfuerzos del estudiante.
  
- 2) La TI tiene flexibilidad para adaptarse a las capacidades y necesidades individuales de cada estudiante: puede trabajar a la velocidad adecuada adaptada a las necesidades de cada estudiante en vez de seguir un ritmo preestablecido. Puede proporcionar diversos puntos de entrada y salida, no juzga e informa al estudiante de su éxito o fracaso sin decir si el estudiante es bueno o malo.

- 3) La TI puede estimular a los estudiantes desmotivados: la información puede ser más interesante, se trabaja con un ritmo propio. Ver en privado que no se ha entendido algo hace que sea más fácil aceptar ayuda.
- 4) Los ordenadores proporcionan a los estudiantes la oportunidad de conseguir buenos resultados donde previamente han fallado.
- 5) Los ordenadores pueden reducir el riesgo de fracaso escolar.
- 6) La TI puede presentar la información mediante formas nuevas que ayuden a comprenderla, asimilarla y utilizarla más eficazmente.
- 7) La tecnología interactiva motiva y estimula el aprendizaje.
- 8) Los programas informáticos con voz digitalizada pueden ayudar en el proceso de lecto-escritura.
- 9) Los ordenadores ayudan a los estudiantes a aprender cuando se utilizan tareas y actividades significativas y bien diseñadas: tienen resultados claros de aprendizaje, se describen de manera clara y sencilla, ofrecen oportunidades para diversos estilos de aprendizaje y son atractivas. Los estudiantes han de saber de qué manera se espera que hagan su trabajo. Han de saber: qué aprenderán, por qué lo han de aprender, si la actividad es individual o en grupo, cuándo tendrán acceso al ordenador y cómo será evaluado su trabajo.
- 10) El uso de la TI hace que los profesores revisen cómo enseñan y cómo aprenden los estudiantes: los estudiantes pueden investigar por sí solos y el profesorado se dedica a proporcionar nueva información, a interrelacionarla con conocimientos adquiridos o redirigir el pensamiento cuando el aprendizaje se bloquea.

## **2.6 Otros aspectos importantes**

Podríamos señalar diferentes aspectos generales que son reconocidos en el aprendizaje de lenguas y que tampoco conviene olvidar en relación con el tema que nos ocupa.

### **2.6.1 La motivación y el interés**

Brett (1999) y Watts (1989) encuentran evidencia en el uso de ese tipo de aplicaciones para trabajo autodidacta por encima del que pueden ofrecer libros, cintas o vídeos.

Skehan (1989) revisa la investigación de la motivación y también de diversos otros factores, tales como la aptitud, estrategias para el aprendizaje de lenguas y factores cognitivos y afectivos, que más tarde los subdivide en: ansiedad, independencia de campo, toma de riesgos, inteligencia y extroversión / introversión.

Bernaus (1992) comenta su punto de vista del proceso de aprendizaje y adquisición de una lengua extranjera diciendo que el lenguaje constituye un proceso único, donde el factor social puede jugar un factor fundamental.

Setter y Jenkins (2005:1) afirman que es a nivel afectivo a través de la manera por la cuál hablamos y, especialmente mediante nuestro acento, donde proyectamos nuestra identidad étnica, social y regional.

En sus referencias a la motivación, Skehan (1989) formula varias hipótesis sobre este fenómeno:

- La motivación intrínseca produce un interés en la tarea del aprendizaje.
- La motivación resultativa incrementa o reduce la motivación según exista éxito o fracaso en una tarea.
- La causa interna, que es la motivación que los alumnos traen con ellos en la situación de aprendizaje.
- La presión externa influenciará la motivación.

Good y Brophy (1987), al referirse a tecnología y aprendizaje, sugieren cinco condiciones iniciales para motivar que deben llevarse a cabo por el profesor como *facilitador* en la clase:

- a) Un apropiado nivel de desafío o de dificultad
- b) Objetivos de aprendizaje que sean significativos para el alumno
- c) Variación en los métodos de uso
- d) Retroalimentación intrínseca (a nivel muy primario, utilizando el procesador de textos, puede dar al alumno un cierto grado de satisfacción) y extrínseca
- e) Inexistencia de barreras para el aprendizaje

Brett (1999) y Crookes y Schmidt (1992) vieron cinco dimensiones de clase que pueden afectar o ser afectadas por factores que están en relación con la motivación:

- La fase preliminar de la clase
- Las actividades de clase
- El *feedback* o retroalimentación
- Los efectos de la percepción de los estudiantes y de los materiales

El hecho de que los alumnos hayan expresado su interés en ciertas áreas, por ejemplo la pronunciación, puede hacer que ciertas actividades en este campo se puedan integrar. Según las conclusiones de la tesis doctoral de Brett (1999) sobre el multimedia como medio, éste no proporciona una motivación a largo plazo. Sin embargo, cuando la motivación para usar la aplicación era extrínseca, las actitudes no fueron positivas. La aplicación no tuvo la habilidad de motivar al aprendiz hasta el punto de usarla cuando no fuera estrictamente necesaria.

### **2.6.2 Multimedia y niveles de motivación**

Situándonos en el término que Brett (1999) designa como un *micro nivel*, el multimedia puede incrementar la atención y ayudar en el desarrollo de estrategias útiles para el aprendizaje de lenguas:

*“Learners’ motivation to deploy useful learning strategies might also be beneficial to acquisition. Multimedia-delivered materials may well be capable of providing the forum for supporting, encouraging, or motivating the use of metacognitive strategies such as “monitoring” and being able to evaluate one’s own performance. Directed attention may be encouraged through meaning-focused tasks with feedback. It might be that in the multimedia environment such metacognitive strategies can be used more easily, or conversely, it might be that it is the multimedia environment itself which motivates the use of such strategies”.*

Brett (1999:14)

Según él, el hecho de que se pueda configurar el entorno multimedia permite que haya una cierta flexibilidad y se pueda acomodar con estudiantes de diversas interlenguas, diferentes estilos cognitivos, diferentes estilos de aprendizaje y diferentes necesidades. He aquí los contenidos que ofrece el multimedia educativo según Brett:

- Aprendizaje del contenido: lo que se puede ver, lo que se puede escuchar y qué cantidad.
- Modo de aprendizaje: vídeo, audio, con o sin subtítulos.
- Activo o pasivo: completar la tarea o no.
- Tipo de tarea.
- Nivel de dificultad en niveles o escalonado.
- Secuencia de aprendizaje: qué hacer, cuándo y en qué secuencia.
- Tiempo y ritmo de aprendizaje.
- Consejo en aprendizaje de lenguas y manera de explotar el programa a través de Internet.
- Materiales del mundo real: a través de vídeos, CD-Roms, audiciones, etc.

### **2.6.3 Condiciones para entornos de aprendizaje de lenguas**

Diversos autores han propuesto diversos modelos de entornos de aprendizaje, como Warschauer (1996), Chun y Plass (2000).

Egbert *et al.* (1999: 3-6) se refieren a una serie de condiciones óptimas de aprendizaje de lenguas:

1. Los alumnos han de tener oportunidades para interactuar y negociar el significado.
2. Los alumnos están involucrados con tareas auténticas.
3. Los alumnos están expuestos a, promocionan el uso de y producen lenguaje variado y creativo.
4. Los alumnos tienen suficiente tiempo y retroalimentación.
5. Los alumnos están guiados para atender conscientemente el proceso de aprendizaje.
6. Los alumnos trabajan en un ambiente con un nivel ideal de stress y ansiedad. Brett (1998a) ha utilizado unos términos similares,

tales como *libre de juicio y libre de peligro*, que se pueden relacionar con este punto.

7. Los alumnos tienen un nivel de autonomía.

#### **2.6.4 Desventajas en el uso de las aplicaciones multimedia**

Brett (1999) se refiere al coste y al equipamiento como el principal problema. El coste de los equipos va bajando progresivamente, pero las novedades y el hecho de que los equipos se vuelvan obsoletos es un grave problema a tratar.

Edwyn James<sup>9</sup>, del Centro para la Innovación y la Investigación de la Educación de la OCDE (Organización para la Cooperación y el Desarrollo Económico) afirmó en el año 2001 que la mayoría de los ordenadores tenían un período de vida de cinco años, en el mejor de los casos. James se preguntaba cómo se podía justificar una inversión de grandes cantidades de dinero en *hardware* en una escuela que sólo abría seis horas al día, cuarenta semanas al año, si en cinco años el material estaría obsoleto. Las posibles soluciones pasaban por permitir el acceso al resto de la comunidad, lo cual facilitaría y rentabilizaría esas inversiones.

Los recursos informáticos y audiovisuales de una escuela pueden no satisfacer a todos. Si bien cada vez más el número de personas interesadas en trabajar con ellos aumenta, hay que trabajar en la formación de profesores a nivel informático para superar los obstáculos con los que se pueden encontrar, como son la solución de pequeños problemas técnicos y la mejora en la búsqueda de sus propios recursos didácticos y planificación escolar.

---

<sup>9</sup> En Fine (2001)

El cansancio en relación a la dificultad de lectura en un ordenador ha sido tratado por Thibodeau (1997), que afirma que es más cansado, más difícil y un 28% más lento el leer en un ordenador que leer en formato papel.

Milton (1997) indica que el problema con la aplicación de materiales en CD-Rom que dependen de becas y ayudas gubernamentales, europeas o privadas es siempre el mismo, la falta de continuidad o el excesivo coste del proyecto. Proyectos como “Tune in English<sup>10</sup>”, cuyo prototipo fue mostrado en unas jornadas de fonética y que contribuía especialmente a la enseñanza de la prosodia, nunca vio la luz quizás debido a estos problemas.

### **2.6.5 Desventajas lingüísticas en diversos estudios sin medios multimedia**

Diversos autores se han referido a problemas muy concretos observados en actividades lingüísticas que podemos relacionar a continuación. En la introducción de su libro sobre la audición, White (1998) enumera una serie de problemas que se producen en la clase de lengua extranjera:

- a) No existe mucho tiempo para escuchar una cinta.
- b) No se dedica mucho tiempo para analizar lo que se contestó mal.
- c) Frecuentemente, la clase de audición hace que los alumnos sean sujetos pasivos.
- d) Las tareas no tratan de vincular la audición con el habla.
- e) A menudo se espera el 10% de la comprensión.

White (1998) cita a Bone (1988), el cual, en un estudio con hablantes nativos, indicó que la gente a menudo escucha un 25% de su potencial e ignora, olvida, distorsiona y malinterpreta el restante 75%.

---

<sup>10</sup> Kaltenboeck (2001)

La concentración aumenta un 25% si ellos consideran que lo que están oyendo es importante y están interesados, pero nunca alcanza el 100%.

## **2.6.6 Los estudios de la memoria**

Un apunte importante en relación a la memoria humana debe considerarse en esta tesis. Si bien este no es uno de los pilares de trabajo y puede abordarse en posteriores tesis, es necesario hablar del procesamiento de la información humana y poder establecer una cierta relación con los métodos multimedia.

### **2.6.6.1 La Psicología cognitiva y los ordenadores**

Navarro Guzmán (1993:173) considera que, con el desarrollo de la Psicología Cognitiva, los procesos mentales vuelven nuevamente a ser objeto de interés de la Psicología científica, debido fundamentalmente al auge de los ordenadores, que analógicamente con el hombre, reciben, almacenan y recuperan información siguiendo estrategias predefinidas. Por esta razón, se produce una incorporación masiva de términos cibernéticos, tales como: almacenamiento, codificación, búsqueda, recuperación y un desplazamiento del enfoque asociacionista del estímulo-respuesta, dando paso a un enfoque cognoscitivo y experimental. Los términos codificación, retención y recuperación se utilizan para definir los tres aspectos básicos de los sistemas de memoria que tienen lugar bajo el control del sujeto. La codificación se refiere a la manera en que se almacena o representa la información; la retención, a la conservación de esta información durante un período de tiempo más o menos prolongado y, la recuperación, a todos aquellos procesos que hacen posible su utilización posterior.

El mismo autor, al referirse a la neurofisiología de la memoria y el aprendizaje, afirma:

*“El aprendizaje es el proceso de adquirir un conocimiento sobre el mundo, es decir, la adquisición a través de la experiencia. La memoria es la retención o almacenaje de ese conocimiento, así como su evocación o consecuencias”.*

Navarro Guzman (1993:51)

Prácticamente no hay actividad nerviosa que no influya o no sea afectada en alguna medida por el aprendizaje y la memoria, ya que la segunda vez que hacemos o percibimos algo, nunca es igual a la primera.

Saettler (1990: 323), en referencia a los ordenadores, citaba a Neisser (1967) diciendo que dichas máquinas proporcionaron a los psicólogos la garantía de que los procesos cognitivos fueran tan reales como las respuestas musculares y glandulares de la conducta humana.

### **2.6.6.2 El modelo de memoria Atkinson-Shiffrin**

Si consideramos uno de los modelos de memoria más aceptados en la actualidad<sup>11</sup>, el modelo Atkinson-Shiffrin, propuesto en 1968, considera que la memoria tiene tres almacenes: los registros sensoriales (incluyendo memorias icónicas y ecoicas) o el almacén de memoria sensorial a corto plazo (STSS), la MCP (almacén de memoria a corto plazo), a veces llamada también memoria de trabajo (WM) y la MLP (a largo plazo). La información en el STSS decae rápidamente (al cabo de de un segundo) y se perdería totalmente si no se transmitiera a la MCP, que tiene una capacidad pequeña y la información decae en un tiempo de 5 a

---

<sup>11</sup> La referencia que se hace hacia el modelo Atkinson-Shiffrin (1968) en relación a uno de los modelos de memoria ha sido la que se ha encontrado como más empleada al consultar diversos autores (Miller, 1956; Peters y O'Sullivan, 2003; French y Anselme, 1999). Otro modelo, como la Teoría de los Niveles de Procesamiento ( Craik y Lockhart, 1972) no se considera como un modelo rival o en desacuerdo con el que aquí definimos.

18 segundos. Una de las características de la MCP es el mecanismo de ensayo, que se ve parcialmente bajo el control del individuo, pero con una capacidad de procesamiento fijo. Ya que la MCP tiene capacidad limitada<sup>12</sup>, la información no transferida directamente a la MLP se pierde.

Navarro Guzmán (1993:191), en relación a la MLP, comenta que el principal problema de este almacén de memoria es cómo recuperar la información que en él se encuentra. El acto de recordar sería, en este sentido, la aplicación sistemática de reglas para analizar esta información almacenada, utilizando estrategias y mecanismos mentales que tienen mucho que ver con el pensamiento y la asociación de problemas. La recuperación va a depender, en gran medida, del papel crucial que tenga la organización en el momento del aprendizaje.

El proceso de memoria sensorial empieza con señales físicas (visuales, acústicas o táctiles) que se reciben por los registros sensores (ojos, oídos y piel). Neisser (1967) ha llamado al input visual como “icono”, al auditivo como “eco” y al tercero como “táctil”.

En relación al modelo Atkinson-Shiffrin y el progreso en la teoría de la memoria, la siguiente cita se refiere a las diversas posibilidades de permanencia de ítems en la memoria:

*“...Un aspecto fundamental en el modelo Atkinson y Shiffrin es que cuanto mayor es la permanencia de un ítem en la MCP, mayor es la probabilidad de que dicho ítem pase a la MLP. Como el único modo de aumentar la permanencia de un ítem en la MCP es repasarlo, cuantas más veces se repase, mayor será la probabilidad de que se transfiera a la MLP. Ahora bien, el paso de un ítem a la MLP no garantiza plenamente su recuperación, puesto que cuando deja de repasarse en la MCP empieza a debilitarse su huella en la MLP”.*

Navarro Guzmán (1993:233)

---

<sup>12</sup>La evidencia sugiere que la MCP es muy limitada y que se ha de activar en ella sólo un pequeño número de ítems de información en un tiempo establecido. En Saettler (1990:325), se sitúan esos ítems en unos siete objetos, o siete unidades de información, y que para retenerlos, la única opción es el ensayo.

### 2.6.6.3 Estudios sobre los efectos de la memoria

En los años 1960, en los laboratorios de teléfono Bell -encabezados por George Sperling- se realizaron diversos estudios visuales y auditivos, utilizando una técnica conocida como *el informe total*, según la cual se puede controlar cuántos estímulos (letras, palabras, dígitos) son captados de un solo 'golpe de vista'.

Unos años más tarde, Darwin *et al.* (1972) también investigaron la memoria auditiva y vieron que la duración temporal del almacenamiento auditivo era generalmente más larga que la del almacenamiento visual.

Penney (1989) recuerda la división del almacenamiento de la información presentada según sea por vía oral o visual al hablar de la retención de material verbal en la memoria a corto plazo, y según su estudio previo de 1975, dice:

*"If there are separate processing streams for auditorily and visually presented items in short-term memory, it should be possible to show that retention is greater if information is presented to two modalities than when only one mode is used".*

Penney (1989:401)

Gathercole *et al.* (1992), en sus estudios sobre la memoria de trabajo, recogen brevemente el estudio de Service (1992), que evaluó las habilidades de repeticiones de palabras que no tenían significado con un grupo de niños finlandeses de nueve años que aprendían inglés como lengua extranjera, considerando los progresos como muy positivos y fueron relacionados dos años y medio después. Este autor argumenta que la asociación entre la repetición de palabras sin significado y los logros de los niños en el aprendizaje de lengua extranjera están mediatizados por la contribución de la memoria de trabajo fonológico en la adquisición de vocabulario.

En relación al tratamiento de la pronunciación, debemos reflexionar acerca de los aspectos descritos anteriormente, ya que no debe tratarse de un recurso aislado y para fines única y exclusivamente orales. El hecho de poder trabajar con elementos multimedia permite integrar magníficamente tanto los aspectos orales como los visuales y establecer relaciones sobre vocales y consonantes que equivalen o se asemejan a sonidos (largos y cortos), dado que la integración del Currículo educativo contempla diversas áreas de interés (comprensión, expresión, orales y escritas).

Brett (1999), en su capítulo sobre el desarrollo del control en el uso del conocimiento de la L2, habla del modelo de procesamiento de la información de McLaughlin (1987). Este modelo tuvo un impacto en la pedagogía del aprendizaje de lenguas. Este autor observa dos maneras de procesar la información en relación con la entrada de datos: el procesamiento controlado y el procesamiento automático. Los dos modos de procesamiento requieren la participación del conocimiento que ya ha sido adquirido y almacenado en la memoria como nodos. Para facilitar la comprensión acerca de lo que es un nodo, Brett (1998b) cita a Shiffrin y Schneider (1977) para referirse a la memoria, diciendo que...:

*“... es una gran colección de nodos que pasan a estar ínter asociados de manera compleja a través del aprendizaje. Cada nodo es un grupo o un conjunto de elementos de información. Muchos de los nodos son interactivos o pasivos y, cuando están en este estado, el sistema interconectado de nodos se llama Memoria a largo plazo”.*

Brett (1998b: capítulo 2.4.4. online)

El segundo modo de procesar la información es el de manera automática; éste ocurre cuando la entrada de datos (*input*) específica de L2 se activa y se conecta con un conjunto de nodos. Este conjunto conectado de nodos se ha formado como resultado del aprendizaje.

El procesamiento automático es el resultado de la práctica utilizando el modo de procesamiento controlado. Gracias a la participación en actividades en tiempo real, los alumnos tienen la oportunidad de utilizar estos procesos cognitivos que automatizan el conocimiento.

#### **2.6.6.4 El papel de la música y la canción en relación a la memoria**

Murphey (1992: 6-7) habla de la importancia de la música y de la canción en el aprendizaje de lenguas. A continuación se citan algunos aspectos que consideramos importantes:

- a) El fenómeno de “la canción *que se repite incesantemente en mi cabeza*” parece reforzar la idea de que las canciones trabajan nuestra memoria a corto y a largo plazo<sup>13</sup>.
- b) La recitación de canciones se parece a lo que Piaget (1923) describió como lenguaje egocéntrico, en el que los niños hablan sin prestar atención a nada en especial. Simplemente disfrutan con repetir. Podría decirse que la necesidad de ese lenguaje egocéntrico nunca nos abandona y se lleva a cabo parcialmente a través de la canción. Krashen (1983) ha sugerido que esta repetición involuntaria puede ser una manifestación de lo que Chomsky denominó “el mecanismo de adquisición de lenguas”. Parece que nuestros cerebros son propensos de una manera natural a repetir lo que oyen en el ambiente para poder darle sentido. Las canciones pueden activar de manera importante el mecanismo de repetición de la adquisición de lenguas, donde también podemos integrar la pronunciación. Ciertamente parece que es así con los niños, que aprenden casi sin esfuerzo.

---

<sup>13</sup>Conceptos citados en Murphey (1992) tomados de su libro anterior (1990b) *The song stuck in my head phenomenon: a melodic din in the LAD? System*18/1, 53-64. Consideramos importante recalcar dichos aspectos, ya que existen prácticas derivadas para trabajar pronunciación en producciones musicales.

#### **2.6.6.5 El ensayo mental en pronunciación. Hallazgos**

Barrera-Pardo (2004), en su estudio relacionado con la precisión, recogió información relevante en tres áreas del aprendizaje de la pronunciación que normalmente escapan a la observación experimental directa. Barrera-Pardo comenta el uso de tres estrategias diferentes:

- 1) La evaluación mental: estrategia usada por el aprendiz que no implica ensayo alguno de lo ya estudiado, sino que simplemente hace reflexionar al alumno sobre lo trabajado.
- 2) El ensayo mental espontáneo: ya tratado en Krashen (1983), que describió el ensayo involuntario de las estructuras de L2 (palabras, sonidos y frases) que hemos tratado en el punto anterior.
- 3) El ensayo encubierto en pronunciación: un tipo de práctica voluntaria en la que el aprendiz practica (repitiendo, recordando o imitando) lo que ha aprendido. Esta estrategia sirve para diferentes propósitos, tales como una herramienta de autocorrección, una manera de sintonizar con material ensayado y también como rastreo del habla al que el aprendiz está expuesto.

Dickerson (1994) conceptualiza el ensayo encubierto de la pronunciación como una estrategia de aprendizaje activa, cognitiva y de auto-evaluación, que requiere cierto esfuerzo e implica un uso consciente de reglas y otros recuerdos de producciones orales, que pueden repararse y practicarse hasta mejorar y sentirse bien. En su experimento con dos grupos (uno con alumnos de pronunciación avanzada y otro con menor nivel en pronunciación), las conclusiones del estudio indicaron que los alumnos de mayor nivel mostraban un uso superior de estrategias, y que

en general practicaban más de manera encubierta sus propias autocorrecciones y accedían a ellas- que los alumnos de menor nivel.

Dichos hallazgos sugieren que la evaluación mental del habla y la atención a la práctica encubierta de la pronunciación pueden ser la característica de los hablantes más avanzados.

## **2.7 El aprendizaje de lenguas asistido con ordenadores (CALL)**

Los diferentes libros que hacían referencia al CALL empezaron a publicarse en la década de los años 1980. Si bien éste es un campo aún inmaduro<sup>14</sup>, el uso de ordenadores en la instrucción del lenguaje está basado en la fusión de grandes tendencias en la educación de los años 50 y 60 del siglo pasado. Es necesario referirse al Audiolingüismo, ya que este método fue una aproximación teórica y metodológica al aprendizaje de lenguas -también con raíces conductistas- que estaba basado en la formación de hábitos y se implementó para ser utilizado en un programa de aprendizaje con un ordenador o también en un laboratorio de idiomas.

Según Larsen-Freeman (1986), el método audio lingual se desarrolló en los EEUU durante la Segunda Guerra Mundial, en un momento en que era necesario aprender una lengua extranjera debido a propósitos militares. Mientras el objetivo de comunicación en la lengua de estudio era el objetivo del método directo, existían en ese momento nuevas ideas acerca del lenguaje y del aprendizaje que emanaban de las disciplinas de

---

<sup>14</sup> Hubbard (1996) dice que el CALL no es un campo maduro, si lo fuera, podríamos definir la metodología CALL como un conjunto de métodos empleados por los profesionales de CALL, donde el método representa un conjunto organizado de principios consistentes y su realización es una parte de un curso educativo o una actividad basada en ordenadores.

la lingüística descriptiva y la psicología conductista. Dichas ideas contribuyeron a desarrollar el método audio lingual.

A nivel tecnológico, la invención de los laboratorios de idiomas prometía la realización de principios de aprendizaje conductistas nuevos y efectivos. El modelo de repetición de ejercicios era el método más usado, practicando la imitación y el refuerzo de la lengua oral y ninguna tolerancia hacia el error.

Lado (1964) resume una serie de principios en los que está basado el método audio lingual y que citaremos a continuación: hablar antes que escribir, frases básicas, modelos convertidos en hábitos, demostración del sistema de sonidos, control del vocabulario, enseñanza de los problemas, la escritura como representación del habla, modelos clasificados, práctica del lenguaje *vs.* traducción, estándar de lengua auténtica, práctica, formas de respuesta, velocidad y estilo, refuerzo inmediato, actitud hacia la cultura de estudio, contenido y finalmente, aprendizaje como resultado crucial. La crítica básica hacia este método fue que se convertía en una práctica mecánica de una lengua dada y que no animaba a los aprendices a crear un uso en aquello que aprendían.

### **2.7.1 Marco metodológico del CALL**

Pennington (1996) opina que los ordenadores ayudan a construir y elaborar las gramáticas de la L2 estructurando el *input* del alumno de diversas formas, por ejemplo, limitando y concentrando el lenguaje y las tareas que se realizarán. El ordenador es un método que responde positivamente y que ofrece a los usuarios una retroalimentación fiable y de alta calidad, generalmente de una manera instantánea. En su tratamiento del marco metodológico del CALL, indica cuatro aspectos a tener en cuenta:

- Jugador 1: alumno
- Jugador 2: promotor (diseña materiales)
- Jugador 3: evaluador (analiza el programa educativo de ordenador)
- Jugador 4: profesor

Una configuración adecuada de los métodos multimedia contribuirá a facilitar el conocimiento de la L2. Gimeno Sanz (2002), en relación al diseño y creación de paquetes multimedia CALL, considera que, una vez se ha tomado la decisión de crear algo, es esencial considerar los siguientes aspectos:

- El grupo de destino
- El nivel de lenguaje del grupo de destino
- El propósito del paquete en sí mismo: para uso general o específico
- El modelo metodológico
- El uso apropiado de la tecnología

Desarrollar CALL requiere una colaboración estrecha entre el profesor de lenguas y el programador. El diseñador gráfico es importante, y un ingeniero de sonido y un técnico de vídeo serán necesarios si el paquete ha de contener cantidades substanciales de sonido e imagen en movimiento.

Levy (1997) ha indicado que el papel del aprendizaje de lenguas asistido por ordenador se puede dividir en dos funciones principales: una como tutor y otra como herramienta. Su encuesta internacional sobre CALL en el año 1990 indicó algunos hallazgos relevantes, en los que la mayor parte de las personas que respondieron describieron ese tipo de aprendizaje como una herramienta. Asimismo, indicaron en ese año que resultaba un obstáculo el trabajar con materiales CALL en desarrollo, ya

que muchos carecían de validez académica y ese desarrollo de programas educativos por ordenador<sup>15</sup> contribuía a una promoción personal o a una prioridad institucional.

Phillips (1986) sugiere el análisis de siete facetas de materiales CALL:

1. Tipo de actividad (*juego, prueba o ejercicio*).
2. Tipo de aprendizaje: énfasis en el aprendizaje.
3. Foco de atención en el aprendiz: en lo que trabaja el alumno mientras interactúa con la lección.
4. Foco de atención en el programa: propósito lingüístico de la actividad.
5. Dificultad del nivel de lenguaje: nivel de lengua usado en el programa.
6. Dificultad del programa en sí: qué nivel de tolerancia tiene el programa con el alumno.
7. Gestión de la clase: cómo se usa el programa y se integra en el conjunto de la clase.

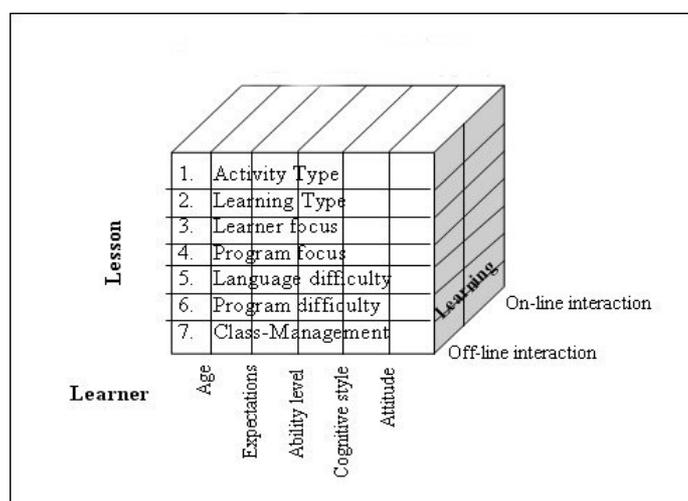


Figura 15. Chapelle y Jamieson, (1991:58). Dimensiones de análisis, alumnos y aprendizaje

<sup>15</sup> Levy también resalta que existen 24 disciplinas y campos de estudio en los cuales los desarrolladores de programas educativos por ordenador y los investigadores creen que han influenciado su trabajo.

Las dimensiones de análisis de lecciones, estudiantes y aprendizaje, según Chapelle y Jamieson (1991:58), quedan resumidas en la figura 15. Las características de los alumnos son importantes para asesorar la efectividad del CALL. Así pues, un programa educativo apropiado puede ser diferente para alumnos de diferentes edades, las expectativas basadas en experiencias de aprendizaje anterior serán importantes, el nivel de habilidad, el estilo cognitivo (independencia de campo / dependencia y reflexión / impulsividad), la relación dinámica entre la actitud y las variables del CALL contribuirán a asesorar su trabajo.

Estebanell *et al.* (2003) y Estebanell (1998) afirman que los alumnos suelen disfrutar si el CALL que usan tiene significado y es apropiado para ellos, e intenta aplicar dichos conceptos con alumnos de Magisterio como se traduce en el tipo de enseñanza de Nuevas Tecnologías aplicadas a la enseñanza. Según esta autora, debemos tratar de reflexionar profundamente acerca de las posibilidades y repercusiones de tipo didáctico de las TIC, generando un bucle permanente entre el conocimiento, el uso y la reflexión. Este proceso debe interiorizarse hasta convertirse en saber personal.

### **2.7.2 La entrada y salida de datos en los medios multimedia**

Brett (1999) argumenta que el subministro de la entrada de datos en L2 es un prerrequisito para SLA, es decir, la adquisición de una segunda lengua, y la tecnología multimedia proporciona una variedad de medios que preven esa entrada de datos: sonido, animación, texto escrito, vídeo y gráficos. El beneficio del multimedia es que puede proporcionar una entrada de datos lingüísticos que simultáneamente transmita la misma información a través de medios diferentes, ya que se dan canales diferentes de procesamiento de información: un canal acústico gracias al

sonido y un canal visual gracias a la animación del vídeo y del texto escrito.

El multimedia necesita proporcionar recursos extra que le permitan recuperar el significado desde la entrada de datos, aumentando este último punto gracias a la ayuda de los subtítulos, los diccionarios y las tareas de comprensión, con las cuales los alumnos tienen las herramientas necesarias para entender esa información. También, esta tecnología necesita que el alumno pueda tener el control para utilizar dichas características. El alumno puede ir a su ritmo gracias a la interacción personal con el ordenador.

Holliday (1999: 182) define la palabra *input* diciendo que en su sentido amplio se puede referir a todo el lenguaje de la lengua que se aprende, a la que el alumno está expuesto y que potencialmente le proporciona un conocimiento sobre ella. Los orígenes de investigación de la entrada de datos desde la perspectiva interaccionista se iniciaron con Ferguson (1971) cuando se refería al concepto de *foreign talk* o habla extranjera, como registro específico de *input* simplificado utilizado por los hablantes nativos cuando hablaban con aprendices de segundas lenguas.

Otros términos que necesitan definición son el *Comprehensible input* (Krashen 1977, 1980, 1985) aplicando el *Monitor Model Comprehensible Output* (Swain, 1985), quien indicó que la habilidad de los aprendices para entender las entradas de datos no garantizaba la adquisición de ciertas formas y estructuras, incluso cuando los aprendices que intentaban procesar el significado de un mensaje, podían no prestar mucha atención a la forma.

Debemos también considerar algunos factores internos en la adquisición de lenguas relacionados con los estudios que hemos citado en relación a la memoria.

Ellis (1997:119) describe el proceso del conocimiento implícito del aprendizaje (ver figura 16) según el cual en el primer estadio, el *input* (entrada de datos) se convierte en *intake* (internalización); éste implica que el aprendiz sea consciente de las características del lenguaje en el *input*, adsorbiendo las entradas de datos en la memoria a corto plazo y comparándolas a las producidas en el *output* o salida de datos.

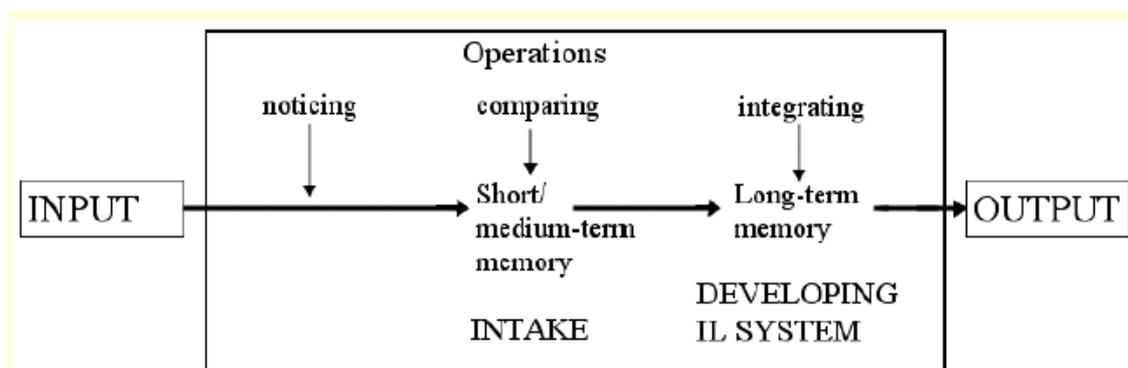


Figura 16. Ellis (1997:119). El proceso del conocimiento implícito del aprendizaje

Otros autores han dado otras definiciones para diferenciar el *input* y el *intake*:

*“Los datos de lenguaje potencialmente procesables y que están disponibles para el aprendiz de lenguas” [input]... “la parte del input que ha sido procesada y convertida en conocimiento de alguna forma [intake]”.*

Sharwood (1993: 167)

También Holliday (1999) cita estudios de Day y Shapson (1991) y Lightbown y Spada (1990) en los que también se ha sugerido que la adquisición de ciertas características lingüísticas puede requerir además *una atención hacia la forma* durante la retroalimentación en la producción de los aprendices.

Actualmente, en la investigación del *input* y del *output* se puede añadir el haber investigado los siguientes puntos:

- a) Los aprendices no están igualmente listos para todos los aspectos de la lengua de estudio al mismo tiempo.
- b) El desarrollo ocurre como resultado de los diferentes niveles de atención en el *input*.
- c) La retroalimentación correctiva en el *output* tiene un rol en el aprendizaje de lenguas, pero no es universalmente la mejor estrategia.

Todos los puntos tratados en relación a las entradas y salidas de datos son aplicables al CALL, ya que los ordenadores influyen el ambiente en el que los aprendices están expuestos a la segunda lengua. Pero no hay muchos programas que presten atención a la interacción en los CD-Roms y programas de *software*. De hecho, muchos de estos programas no interactúan en absoluto y todas las respuestas que puede dar un ordenador están basadas en bancos de datos pre-programados y en audio y vídeo pre-grabado.

Aunque el ordenador no entiende el lenguaje natural, es capaz de recibir (*input*) y producir (*output*) múltiples cantidades de datos. Sin embargo, para que este aspecto sea más viable, debe tener la capacidad de interactuar con el usuario aproximándose a la comunicación real, si bien existen aspectos no directamente relacionados con la mejora en la enseñanza de lenguas.<sup>16</sup>

Brett (1999) comenta que el uso de la entrada de datos bimodal (canal visual y auditivo) puede facilitar y permitir que los alumnos trabajen la

---

<sup>16</sup> Meredith (1978) llevó a cabo un estudio que asesoraba la impulsividad / reflexión de los alumnos y encontró que los alumnos impulsivos realizaban mejor una frase oral (en español) cuando el programa CALL les forzaba a esperar antes de responder.

fonología. También señala que hay cuatro caminos en los cuales el multimedia puede ayudar a trabajar las entradas de datos:

- La previsión de tareas en pantalla designadas para ofrecer especial atención a características lingüísticas.
- La existencia de una retroalimentación<sup>17</sup> para estas tareas.
- Algo que acabemos de resaltar o pronunciar.
- El abastecimiento de recursos que los alumnos pueden usar para negociar la comprensión.

Los medios multimedia necesitan repetir las condiciones que se cree que desarrollan el control sobre el conocimiento de la L2.

## 2.8 Los laboratorios de idiomas

En el primer capítulo, al referirnos a los aprendizajes conductistas y el modelo de repetición de ejercicios, ya se ha introducido muy brevemente qué tipo de herramienta de uso era el laboratorio de idiomas y también se ha introducido el laboratorio portátil ofrecido por el *Departament d'Educació*. En las sucesivas líneas vamos a ver las diferentes posibilidades que ofrece esta herramienta desde un punto de vista global. Para ello, es necesario primero ver la evolución de esta maquinaria, hacer una referencia a qué países la utilizan de una manera más regular y con más avances. Posteriormente, vamos a distinguir entre los laboratorios antiguos, que empleaban un sistema analógico, y los laboratorios actuales, que son digitales.

---

<sup>17</sup> Brandl (1995) investigó las preferencias en el tipo de *feedback* en un programa alemán de CALL, indicando que los tipos de *feedback* eran: Respuesta correcta o incorrecta (1), Descripción gramatical de la respuesta correcta (2) y localización del error y respuesta correcta (3). Los alumnos de nivel alto optaron por el primer y tercer tipo, mientras que los alumnos de nivel bajo preferían la descripción gramatical correcta y visualizar la respuesta correcta.

### **2.8.1 Los inicios de los laboratorios de idiomas**

Según Davies *et al.* (2005), la tecnología digital se remonta a la invención del ordenador digital en Manchester en 1948, si bien hasta mediados de la década de los años 70 no aparecen los primeros micro ordenadores a precios razonables. La inclusión del multimedia dentro del laboratorio digital multiplica las diversas formas de difusión de contenidos, siendo éstas la radio, la televisión digital o terrestre, la grabadora de sonidos, el reproductor de CD o DVD, la Internet, el teléfono móvil o las agendas electrónicas (PDA, siglas en inglés de *Personal digital assistant*). El denominador común en estos aparatos es que la mayoría pueden reproducir vídeo - al cual se refiere Hill (1999) diciendo que es posible estudiar el movimiento de los labios, los gestos relativos propios del lenguaje oral- y por supuesto audio.

Davies *et al.* (2005) se refieren a los laboratorios de idiomas diciendo que fueron populares en escuelas de educación secundaria y otras instituciones a finales de la década de los años 60 y principios de los 70 del siglo anterior. Fueron inicialmente percibidos como una solución al problema del aprendizaje de una lengua a corto plazo para un número elevado de alumnos y se consideró que el invento valió la pena. Sin embargo, su uso decreció a finales de la década de 1970, en parte debido a la falta de confianza en la tecnología, pero también porque se asoció estrechamente con un modelo conductista pasado de moda para la enseñanza y aprendizaje de lenguas. Las razones por las cuales no se obtuvieron los resultados esperados fueron: la falta de un entrenamiento adecuado de los profesores que deseaban emplearlo y la falta de creatividad en proponer actividades más que ejercicios mecánicos; dos aspectos que, en la práctica, aún han de mejorarse en la actualidad.

Stern (1992:10) remarca que la tecnología se convirtió en un aspecto central del nuevo método audio lingual y el laboratorio formaba parte de una nueva era. Igualmente, según Underhill (1989), refiriéndose a esta

herramienta y refiriéndonos a la mejora de la pronunciación, las máquinas probaron ser el objetivo primordial para ayudar al profesor mediante un número inacabable de ejercicios, un modelo de pronunciación perfecto y una manera de liberar al profesor de actividades intelectuales en la clase. Rivers (1990:274), al referirse a la desaparición del método audio lingual y a la potenciación del método comunicativo, comenta que su uso pareció irrelevante y aburrido.

Celce-Murcia *et al.* (1996:311) se refieren a la cinta de casete y la grabadora como herramienta dentro del laboratorio de idiomas, indicando que fueron el medio y el lugar adecuado para que se produjera ese aprendizaje. Con la desmitificación de la psicología conductista y el desencanto resultante en el campo de la lingüística con el método audio lingual, el marco de los laboratorios de idiomas y la tecnología instruccional cayó en desgracia. De hecho, esto ocurrió no tanto porque la tecnología en sí tuviera la culpa, sino porque los métodos y los materiales estaban pasados de moda, tenían limitaciones de aplicación y eran aburridos o carentes de estímulo. Dichos autores comentan que el renacimiento del laboratorio de lenguas representa un triunfo de la tecnología sobre el método y proporciona un testimonio al papel que la tecnología instruccional puede jugar en la enseñanza de lenguas.

A mediados de la década de 1980, el laboratorio volvió a tener un nuevo sentido, gracias en parte a las mejoras tecnológicas, pero también debido a unas mejoras en las herramientas de control, al hacerlas más manejables y sencillas, unos materiales más imaginativos y un mejor diseño del propio laboratorio. Al mismo tiempo, el autoaprendizaje se puso de moda y se produjo una riqueza de nuevas ideas tales como el uso del trabajo en pareja, el empleo de técnicas de situación (*roleplays*), juegos comunicativos, etc., descritos por Ely (1984).

## 2.8.2 La evolución hasta la fecha actual

Hoy en día, las grabaciones de audio y el marco del laboratorio de idiomas continúan siendo valorados para la práctica de la pronunciación, particularmente debido a la aparición en el mercado comercial de programas auxiliares.

Murphy (1993), al referirse a los laboratorios de idiomas, utiliza el término LARC para designar el *Language Acquisition Resource Center*, en una cita que engloba información interesante relativa a la correcta utilización de este recurso informático y a su uso comunicativo en entornos de aprendizaje:

*"If used wisely, a language laboratory is an example of a modern technological resource that can provide a rich source of learning opportunities for ESL learners. Such opportunities can be provided within whole-class instructional settings and through activities that aim to facilitate peer collaborations as well as self-directed learning. However, it will probably fall to L2 classroom teachers, working in cooperation with LARC personnel and program administrators, to guide ESL students in the use of audio and video material, computer programs, and other technological resources through procedures reflective of the current paradigm for L2 instruction".*

Murphy (1993:14)

Dada esta disponibilidad de material auténtico, el laboratorio de lenguas puede proporcionar acceso casi ilimitado al discurso nativo y a todas sus variantes (dialectos, registros, sexo, etc.). Además, el laboratorio puede proporcionar a los alumnos otras condiciones deseables, como son una mayor práctica que no es posible en el ambiente típico de clase, un ambiente desinhibido para imitar sonidos, la posibilidad de comparar su voz con la del modelo y el control de esos materiales, de sus secuencias y del ritmo.

Ya en la década de los años 90, Beauvois (1996) realizó un estudio en el cual se comparaba la producción del alumno en dos ambientes y

concluyó que la conversación generada en el laboratorio de lenguas fue superior al de los debates de clase sobre los mismos tópicos, tanto en calidad como en cantidad.

### **2.8.3 El mercado de laboratorios americano**

Actualmente, en el mercado de laboratorios portátiles existen diversos productos que desafortunadamente suelen ser sólo accesibles, por su coste, en América del Norte (Estados Unidos y Canadá), países pioneros en este tipo de tecnología.

A tenor de los diferentes productos analizados, podemos afirmar que suele existir un denominador común de recursos: una consola para el profesor con la que se ofrecen sesiones de entrenamiento a los profesores que se encargan de utilizar estos productos, un ordenador ‘servidor’ en el que los recursos están almacenados y se ofrecen a los alumnos y auriculares con micrófonos incorporados (normalmente a prueba de malas manos, es decir, diseñados para resistir y ser utilizados muchas horas). Algunas empresas como *CAN 8 Virtual Lab* empezaron la distribución de estos productos en Canadá en 1967, contando también con la ayuda del *National Research Council of Canada* (NRC). *Sanako*, una empresa americana multinacional en colaboración actualmente con la República Popular China y Finlandia, introdujo el primer laboratorio en el año 1961<sup>18</sup>.

---

<sup>18</sup> [documento en línea] <http://www.sanako.com>



Figura 17. Laboratorio de idiomas en EE.UU. Sistema de ahorro de espacio y almacenaje en techo

Algunos productos norteamericanos (por ejemplo *Educational Media, LLC*) apuestan por la reducción o adaptación de espacio e introducen cajas plastificadas colgadas en el techo del aula (ver figura 17). Mediante un mecanismo, bajan a una altura estándar donde el alumno simplemente se ha de colocar los auriculares, activar un micrófono y ponerse a trabajar, en este caso en formato papel. Es el profesor el que, mediante una consola, interactúa con los alumnos. Éstos no manipulan ninguna máquina, con lo cual el riesgo de deterioro disminuye.

#### **2.8.4 El sistema analógico y el sistema digital**

Renschel (2004) comenta que en 1999, un 10% de los laboratorios vendidos en EEUU eran digitales, en el 2002 ya fueron un 85%, y un 99% en 2004. A finales de los años 90 y a principios del 2000 se introdujeron los laboratorios híbridos, de transmisión analógica y almacenamiento digital.

Las razones por las cuales se empleaba un sistema analógico fueron la existencia de una amplia biblioteca de materiales en dicho formato (cilindros, discos, cinta magnética), la limitación de fondos, la saturación de reproductores de casete y, finalmente, la tecnofobia o miedo al uso de tecnología más avanzada.

El hecho de utilizar los productos digitales -entiéndase discos duros, CD-Roms, memorias portátiles- supone la irrupción tecnológica por la cual no sólo se trabaja con audio, sino con medios multimedia. ¿Qué la hace diferente? Existen varios factores: la habilidad de poder realizar la grabación comparativa, la posibilidad de reproducir la pista del programa mientras simultáneamente graba al alumno, así como el permitir al estudiante trabajar a su propio ritmo. Ahora bien, el próximo apartado nos lleva a considerar cuál es el producto más interesante a tener en cuenta.

## **2.9 Tipos de laboratorios**

El laboratorio digital está ligado al uso adecuado de *software* que proporcione la integración de vídeo, un procesador de palabras y aplicaciones informáticas. Se pueden distinguir dos tipos: los que combinan sólo *software* y los llamados híbridos.

El primer tipo es el típico ordenador que solemos tener en casa con la diferencia de que existe una comunicación con el profesor utilizando un controlador especial, mientras que el segundo tipo ya requiere un cableado mucho más profesional e interfaces para combinar la ayuda de la comunicación verbal y las señales de control. Conceptos tales como la versatilidad, la facilidad de movimiento entre diversas aplicaciones, la interactividad, el potencial para la intervención del profesor y el aprendizaje independiente son características básicas en los laboratorios digitales.

En la actualidad, el uso de los laboratorios, sean o no portátiles, depende de varios factores:

- 1) De la prioridad de uso del profesor, si lo cree conveniente.
- 2) De la existencia de aulas de idioma que faciliten el uso de esa tecnología, ya que la mayoría de docentes han de cambiar de aula y no existen muchas salas destinadas al aprendizaje de una LE.
- 3) La utilidad que se le pueda dar a la máquina, así como su máximo rendimiento, no sólo como reproductor, sino como complemento de grabación, uso de materiales exclusivos, uso del micrófono y de los auriculares.

Un componente importante como la motivación puede tener una incidencia distinta dependiendo de su uso. Hunter (2001b), ver figura 18, en la siguiente página, considera que el control del uso por parte del alumno hace éste tenga un alto nivel de motivación, mientras que, el que lo dirija el profesor y que se vincule el uso remedial como mejora de varios aspectos del habla, hace decrecer dicho factor. La parte inferior del gráfico de la siguiente figura explica la diferencia entre el laboratorio CALL y el laboratorio para uso autodidáctico basado en CALL. Las diferencias principales residen en que, si es el alumno quien controla su uso y su mejora, los programas educativos pueden ser muchos, así como su ámbito de uso: auto-mejora, trabajo remedial y deberes.

Para centrar la cuestión, conviene mencionar los aspectos que Hunter incluye en el uso del laboratorio: todo lo relacionado con texto e hipertexto, imagen, sonido, multimedia, vídeo y animación, juegos y funciones de búsqueda en Internet; elementos que no totalmente están relacionados con el uso que intentamos comentar aquí, es decir, el de la integración de una interacción entre la máquina y el alumno, con audiciones y repeticiones de producciones orales.

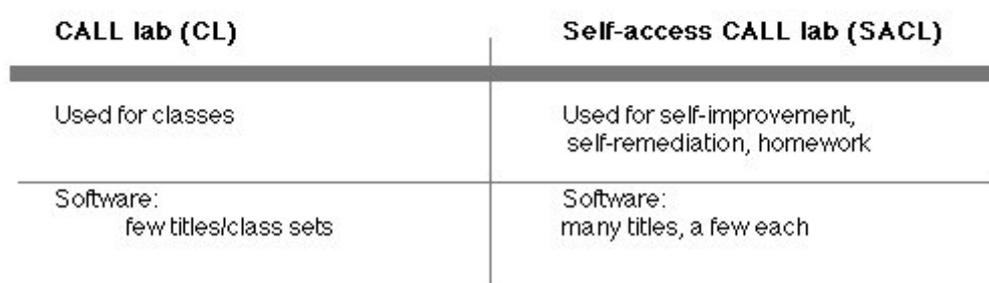


Figura 18. Hunter (2001b). La motivación y el uso de los laboratorios <http://www.info.kochi-tech.ac.jp/lawrie/CALLL/index.html>

Naiman (1992), al referirse al laboratorio de lenguas, incide en que no se necesita un laboratorio caro para enseñar pronunciación, ya que el hecho de que los alumnos puedan grabarse en cintas para reproducir y controlar sus producciones orales en clase así como grabar conversaciones naturales hace que ninguna de estas actividades requiera un laboratorio. De todas maneras, las siguientes páginas van a mostrar ejemplos de tecnología que pueden llevar a cabo una amplísima variedad de interacción entre alumnos.

Gimeno Sanz (2002) apuesta por utilizar actividades del laboratorio de idiomas con alumnos para registrar sus voces y compararlas con un modelo pre establecido, que abarque desde palabras aisladas y frases a unidades más largas de significado, como intercambios en un diálogo.

Remschel (2004) se refiere a los cuatro principios que se consideran indispensables a la hora de referirse a la tecnología digital.

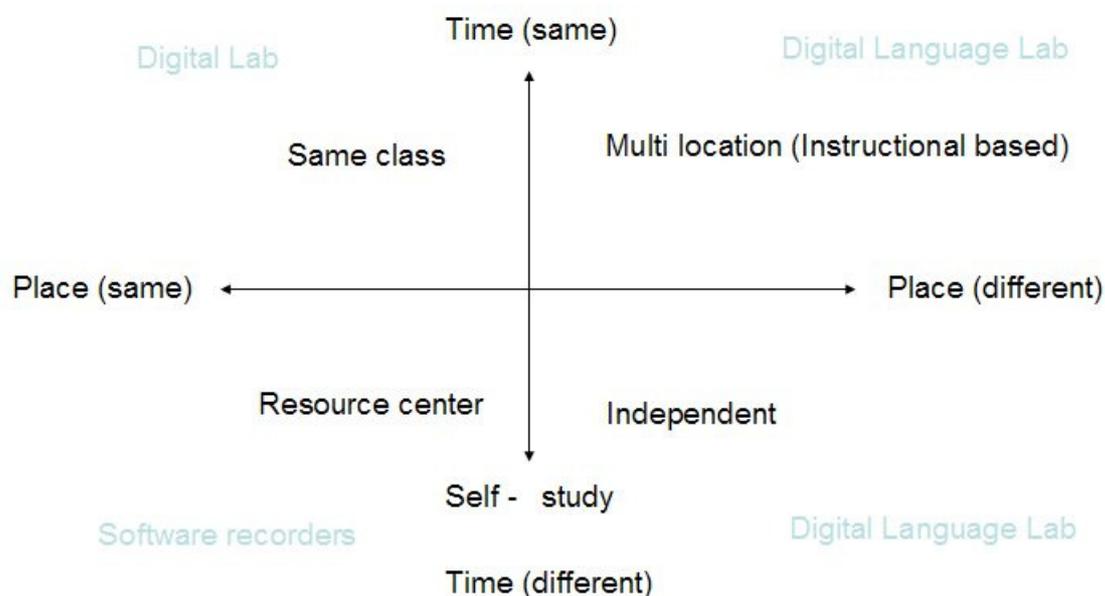


Figura 19. Remschel (2004). “Digital Technology: where we are, where we are going”

En la figura 19, Remschel (2004) divide el uso de la tecnología refiriéndose a su utilización en un lugar u otro, a la vez que sea de un producto u otro, o en un momento diferente. Según estas variables, los productos se pueden englobar para ser utilizados en el aula de idiomas tradicional, en el centro de recursos, en un lugar independiente o en diferentes espacios. El uso del laboratorio digital permite trabajar a la vez o en un momento diferente e inclina la balanza hacia el estudio personal.

## 2.10 El reconocimiento automático de voz. Análisis teórico

En CALL, la señal analógica se convierte en digital y ésta se refleja en una secuencia lógica de declaraciones, expresadas con los números binarios 0 y 1. El desarrollo de tecnología sofisticada, como el ordenador digital, ha introducido la creación de programas de *software* capaces de

extraer el tono, la frecuencia y la intensidad del habla grabada de una manera rápida y precisa, así como la visualización de los modelos acústicos y espectrográficos en una pantalla de un ordenador. El proceso de digitalización de una señal analógica para que pueda ser usada con *software* se resume de la siguiente manera:

- 1) El habla se graba vía micrófono en una cinta de forma analógica.
- 2) Ésta se convierte en digital con un convertidor del ordenador.
- 3) El ordenador la procesa.
- 4) Los resultados se pueden enviar a un monitor, impresora, etc.

La síntesis fonémica es un proceso por el cual cada fonema, o sonido, de un lenguaje está representado por un código digital o filas de símbolos. Este código se traduce por el sintonizador en modelos particulares de frecuencia, duración y amplitud, que modelan los sonidos del habla.

Ehsani y Knodt (1998) revisaron el estado del reconocimiento de voz y de la tecnología con unas aplicaciones en concreto y sugirieron cómo crear entornos de aprendizaje robustos. Según ellos, el CALL ha emergido como una alternativa tentadora a las maneras tradicionales de suplementar o reemplazar la interacción directa profesor-alumno, con los laboratorios de idiomas y el estudio personal basado en cintas de audio. Dichos autores relacionan diversa literatura en relación al aprendizaje asistido por ordenador y llegan a la conclusión de que falta un marco teórico unificado para el diseño y la evaluación de los sistemas con CALL (Chapelle 1997, Hubbard 1988), así como una ausencia de evidencia empírica conclusiva para los beneficios pedagógicos del aprendizaje de lenguas con ordenadores (Chapelle 1997, Dunkel 1991 y

Salaberry 1996). La siguiente cita resume los aspectos explicados anteriormente:

*“Los humanos y las máquinas procesan la información de una manera diferente fundamentalmente (Bernstein y Franco, 1996). Los procesos cognitivos complejos consideran la habilidad humana de asociar señales acústicas con significados e intenciones. Para un ordenador, por otra parte, el habla es esencialmente una serie de valores digitales. Sin embargo, a pesar de esas diferencias, el problema clave del reconocimiento del habla es el mismo para los humanos y para las máquinas: concretamente el encontrar la mejor equivalencia entre una muestra de sonido dado y la correspondiente palabra”.*

Ehsani y Knodt (1998:47)

### **2.10.1 Origen y definiciones**

O’Shaughnessy define la comunicación oral de la siguiente manera:

*“The transfer of information from one person to another via speech, which consists of variations in pressure coming from the mouth of a speaker”.*

O’Shaughnessy (1987:1)

Faudez (2000) comenta los fundamentos del tratamiento del habla y cita:

*“La generación del habla por parte de las personas, consiste en la creación de una onda de presión sonora que se propaga a través del aire a una velocidad de unos 340 metros por segundo. La concatenación de unos determinados sonidos y en un orden prefijado, característicos de cada idioma, constituyen el mensaje”.*

Faudez (2000:32)

La investigación moderna del habla se inició en los años 1930, cuando se desarrolló un método de transmisión práctica digital denominado PCM (*Pulse Code Modulation*). La invención del espectrógrafo en 1946 amplió mucho la capacidad de análisis del habla. No fue hasta los tempranos

años 70 del siglo pasado cuando el reconocimiento del habla explotó su trabajo.

### **2.10.2 Las técnicas del reconocimiento de habla**

Por otra parte, Llisterri (2001:16) define las técnicas de reconocimiento de habla como una comparación, en los métodos más sencillos, entre los patrones acústicos previamente almacenados en el sistema y los enunciados producidos por el hablante.

Varios autores como Stevens (1998), Ladefoged (1996) y Kent y Read (1992) se han referido al análisis de los términos comentados a continuación. En este punto, conviene tener clara una nomenclatura y unos términos que son necesarios para profundizar en los aspectos tratados en esta tesis. También son de gran ayuda las representaciones visuales y comentarios aparecidos en Llisterri (2006) acerca de elementos que veremos a continuación.

## **2.11 Elementos importantes al analizar el habla**

Existen una serie de elementos que conviene mencionar y analizar su funcionamiento para un mejor conocimiento de cómo trabaja el reconocimiento de voz.

### **2.11.1 El espectrograma**

Uno de los elementos esenciales para analizar el habla es el espectrograma -también llamado monograma- que convierte una forma de ondas del habla en dos dimensiones (amplitud y tiempo) en un modelo tri-dimensional (amplitud, frecuencia y tiempo). Con el tiempo y la

frecuencia en el análisis horizontal y vertical respectivamente, la amplitud se mide por la oscuridad del visualizador. Los sonidos sonoros son los que se marcan en el espectrograma, debido a un incremento de la amplitud del habla cada vez que la vocal se acerca. Estos sonidos sonoros son las vocales (incluidos diptongos), excepto cuando se susurran. Los sonidos sordos causan unos modelos rectangulares oscuros puntuados con puntos claros debido a las variaciones instantáneas de energía.

Komissarchik *et al.* (2000) comentan que todas las representaciones reciben el nombre de espectrogramas, o representaciones de la amplitud en función de la frecuencia. La señal de voz es limitada en banda, a unos 8 kHz. Sin embargo, la mayor parte de la información se encuentra en los primeros 4 kHz, que es aproximadamente el ancho de banda utilizado en las comunicaciones por vía telefónica. En relación a la prosodia, el habla conversacional emplea 150-250 palabras por minuto, incluyendo pausas, que promedian unos 650 ms (milisegundos) cada una. Por poner algunos ejemplos, la duración típica de una sílaba es de 200 ms, con vocales acentuadas, éstas promedian 130 ms y la duración de otros elementos fónicos alrededor de 70 ms. Así pues, la vocal /ə/ (schwa) es típicamente de 45 ms, mientras que los diptongos en conversación promedian 180 ms. El promedio vocálico es 30 ms más largo que las consonantes sonorantes, que son 5 ms más largas que las obstruentes. Los diptongos son 75 ms más largos que las vocales. Mediante un micrófono es posible captar la onda de presión sonora emitida por un orador y convertirla en señal eléctrica, digitalizada y poder tratarla en un ordenador.

Lambacher (1996b), al definir las propiedades de los sonidos, se refiere a la frecuencia de un sonido diciendo que éste está determinado por la variación de la presión de aire o el número de ciclos completados por segundo de una onda sonora (medidos en hercios) y, dependiendo de los sonidos, éstos tienen diferentes concentraciones de energía. Por poner

un ejemplo, el sonido fricativo /s/ en inglés, tiene una concentración de energía que oscila entre 4000 y 7000 hz, mientras que el sonido nasal /m/ tiene una baja concentración de energía, entre 250 y 300 hz.

Otro aspecto destacable de los espectrogramas correspondientes a fonemas sonoros es la presencia de zonas enfatizadas, o resonancias, y zonas desenfáticas, o también llamadas antirresonancias. Estas zonas se deben a la configuración física de las diferentes cavidades del tracto configurado por el sistema fonador<sup>19</sup>.

Los formantes aparecen como franjas horizontales, mientras que los valores de amplitud en función de la frecuencia se representan en tonalidades de grises en sentido vertical. Existen dos tipos de señales sonoras que se pueden visualizar: la de banda ancha, en la cual se ven los picos de voz, y la de banda estrecha, en la que se ven unas impresiones de marcas. Dichas señales dependen del ancho de banda del filtro que se haya utilizado para realizar el análisis frecuencial.

### **2.11.2 El osciloscopio**

Un osciloscopio mide la evolución temporal de la amplitud del sonido. Con él es posible obtener una imagen que representa la amplitud de la señal en función del tiempo, que suele denominarse oscilograma. Es importante destacar que tanto el micrófono usado como la sala en la que se realiza la grabación pueden afectar a las tasas de reconocimiento. Llisterri (2001:6) ha tratado de comentar los diferentes datos que se observan en las producciones del alumno en un osciloscopio, así como también la producción nativa o del profesor.

---

<sup>19</sup> Faudez (2000) nombra como componentes principales del sistema fonador humano a los pulmones, la tráquea, la laringe, la cavidad de la faringe, la cavidad oral y bucal y la cavidad nasal. El tracto vocal empieza a la salida de la laringe y termina a la entrada de los labios. El tracto nasal empieza en el paladar y termina en los orificios nasales. Los parámetros principales del sistema articulatorio son las cuerdas vocales, el paladar, la lengua, los dientes, los labios y las mandíbulas.

Más adelante, al analizar diversos programas de pronunciación, podremos profundizar en este aspecto.

Petek (1999) hace referencia a la necesidad de una gran fiabilidad en el interfaz de reconocimiento de voz:

*“In general, multimedia courseware can be considered as an efficient complement to the traditional ways of teaching. Interactivity of the courseware can balance between the teacher guided and open ended learning environments. It can potentially create and stimulate the learners’ motivation and interest. Therefore, it is important that the interactive multimedia courseware offers an efficient user interface”.*

Petek (1999, documento *online*)

### **2.11.3 Otros términos y consideraciones importantes**

Utilizando los principios que contribuyen al éxito en el entrenamiento de la pronunciación, Kenworthy (1987) presenta una lista de cinco principios que se han de aplicar en el entrenamiento de la ASR (*Automatic Speech Recognition*) o reconocimiento automático del habla, divididos en cuatro factores que se refieren al ambiente externo:

- a) Deben producir grandes cantidades de frases por sí mismos. La tutorización individual con instrucción humana suele ser demasiado costosa e impráctica. El tener que compartir la atención del profesor con otros reduce la cantidad de tiempo que cada alumno emplea en crear producciones en lengua extranjera.
- b) Deben recibir la corrección oportuna, para ello los profesores deben asegurarse de que hayan enseñado a los alumnos a producir – articular el sonido correctamente, más que a producir pares mínimos (*minimal pairs*).

- c) Deben oír diferentes modelos nativos, gracias al multimedia ello es posible.
- d) Se debe enfatizar la prosodia (amplitud, duración y tono).

Y uno al ambiente interno:

- e) Deben estar relajados en la situación de aprendizaje. Cuando se fuerza al alumno a producir sonidos que no existen en su lengua nativa delante de sus compañeros, tiende a perder auto confianza.

La ASR se puede usar en la enseñanza de la pronunciación en L2 como complemento del profesor humano. La máquina se hace cargo de esos aspectos de pronunciación práctica que corresponden a los principios anteriormente explicados.

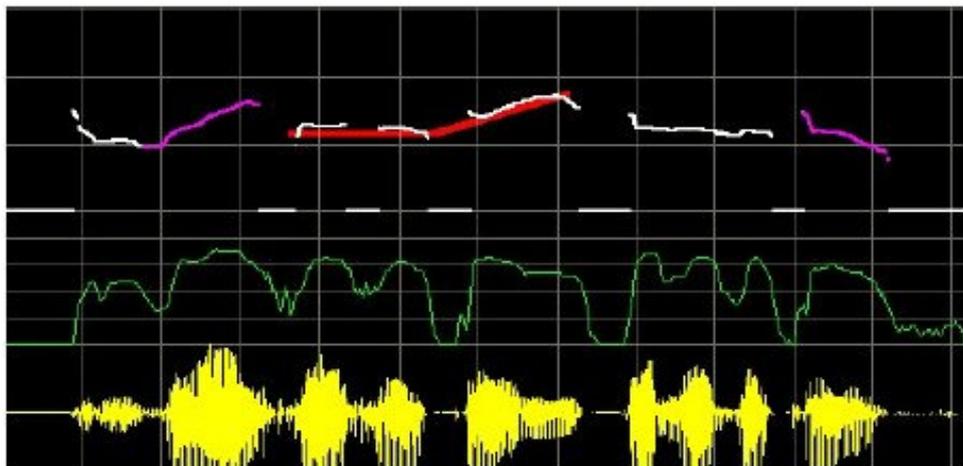


Figura 20. Llisterri (2001:6). Oscilograma, curva de intensidad y curva melódica con el programa WinPitch LTL (Germain y Martín, 2000, figura 2)

Los programas de reconocimiento de voz tratan de llevar a cabo los principios anteriores, algunos con mayor acierto que otros. En dichos mecanismos tratados en los dos puntos anteriores es posible observar la

llamada curva melódica (que se puede observar en la parte superior de la figura 20), es decir, la variación temporal de la frecuencia fundamental y la curva de intensidad (parte central de la misma figura 20) o evolución de la intensidad en el tiempo. Esta figura nos muestra dichos fenómenos así como un oscilograma (parte inferior del gráfico).

La figura 21, a continuación, nos muestra un dibujo que corresponde a un fragmento de un diptongo pronunciado por un hablante masculino.

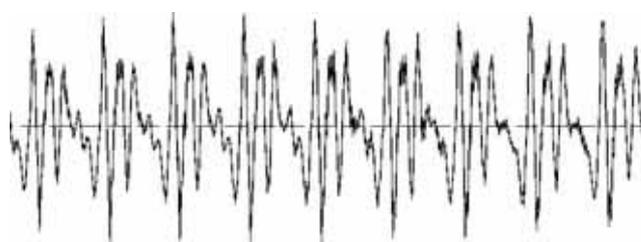


Figura 21. Komissarchik *et al.* (2000:2).Diptongo 'aw' dentro de la frase "How are you?".  
Pronunciación masculina

Komissarchik *et al.* (2000) comenta, en referencia a la anterior figura, que los topes representan la quema de energía que se produce después de que las cuerdas vocales se sacudan y el aire pase a través de ellas. Estos topes se llaman también puntos de excitación. Si una persona produce una excitación durante un segundo, decimos que su tono es de 100 Hz (herzios). El alcance de tono para hombres es de 80-120Hz, para mujeres de 180-220 Hz, y para niños puede llegar a ser de 400-500Hz<sup>20</sup>.

Pellegrino *et al.* (2001) nos ofrece una página web llamada *Ciberphone* a un nivel muy visual y auditivo que puede servir para mejorar la comprensión de ciertos aspectos técnicos descritos aquí de manera más breve.

<sup>20</sup> Si bien Faudez (2000) tiene unos parámetros superiores, diciendo que el margen habitual para locutores masculinos adultos del valor del tono (*pitch*) es de 50 a 250 Hz (período de 20 a 4 ms o 160 a 32 muestras al trabajar con frecuencias de muestreo  $F_m = 8$  kHz, mientras que para locutoras se encuentra entre 120 y 500 Hz (período de 8,3 a 2 ms o 66 a 16 muestras con  $F_m = 8$  kHz).

Dicha página web dividida en tres partes: una introducción a la naturaleza de los sonidos y a su propagación, la descripción de las ondas sonoras y las tipologías de sonidos, y finalmente, las nociones espectrales y los espectrogramas.

En la figura 22, a continuación, que se refiere a un conversor de texto a voz, Faudez (2000) afirma que en los sistemas de conversión texto-voz no es totalmente necesario obtener un excelente grado de inteligibilidad segmental. Según él, la inteligibilidad en la ASR está relacionada con la facilidad para comprender la señal oral. Normalmente, se acostumbra a relacionar la inteligibilidad con el proceso segmental. La calidad es un indicador de la “naturalidad” de los sonidos.

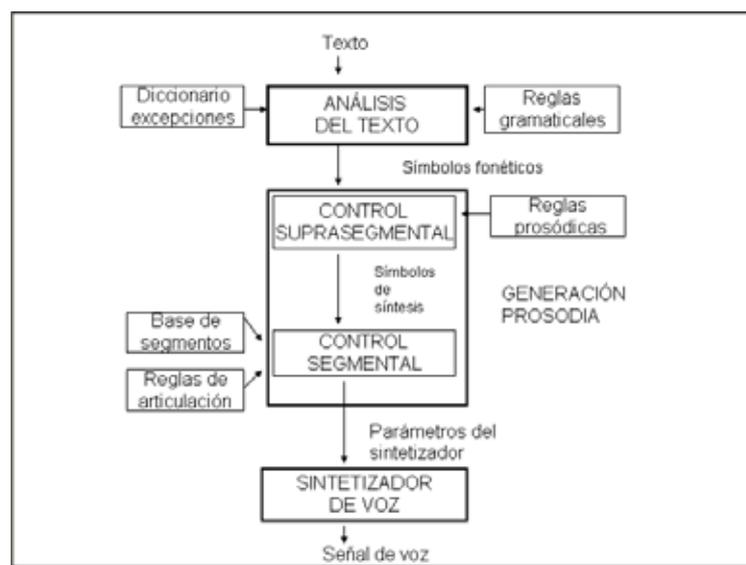


Figura 22. Faudez (2000:82). Conversor de texto a voz

Por lo que hace referencia al procesado prosódico o suprasegmental (conjunto de factores que afectan la pronunciación de manera global, como la entonación, el ritmo y la intensidad del habla), si el texto es relativamente largo, es muy importante que la voz sintetizada tenga una buena entonación y ritmo, y que sea agradable escucharla. Se puede considerar que un convertidor texto-voz consta de tres grandes bloques:

- Análisis lingüístico del texto
- Generación de la prosodia
- Síntesis de la señal de voz

La facilidad para realizar la transcripción fonética depende del idioma que se está tratando. Por ejemplo, resultará más sencillo en español que en inglés, ya que en el primer caso existe una gran concordancia entre la grafía y los fonemas, la cual simplifica el análisis. En cambio, en el inglés, hay que recurrir a un análisis morfológico y existe una complejidad de reglas que ha originado otras alternativas como los modelos conexionistas (estándares fonéticos europeos con caracteres ASCII).

El análisis de la estructura sintáctica permite a los módulos posteriores de generación prosódica determinar correctamente la localización de los grupos de entonación y, para resolver las ambigüedades en la transcripción de los homónimos (es decir, las palabras que en inglés se escriben igual, pero pueden tener pronunciaciones distintas), se hace necesario un análisis semántico del texto que ayude a determinar la pronunciación correcta.

## **2.12 Los diferentes tipos de reconocimiento de voz**

Según Wachowicz *et al.* (1999:255), el CALL de habla interactiva utiliza una variedad de reconocedores del habla y distingue cuatro tipos:

- a) Los sistemas de ASR de hablante independiente, que intentan reconocer casi cualquier hablante de inglés sin tener que entrenar el reconocimiento de voz en el tono o en el modelo de habla de un hablante individual. Estos sistemas son más apropiados en

aplicaciones de aprendizaje de lenguas, especialmente para hablantes no nativos. Estos sistemas han sido diseñados para usar varios tipos de voz y acentos regionales, así que se supone que aceptan todo tipo de hablantes nativos así como entrada de datos comprensibles y no-nativos.

- b) Los sistemas ASR de hablante dependiente, en que el hablante ha de entrenar su voz, y son adecuados para dar órdenes y poder utilizar una computadora.
- c) Los sistemas de ASR de voz discreta, cuyo propósito es reconocer palabras individuales y cuya entrada de datos se ha de separar por pausas. No serían adecuados para tareas a nivel frase.
- d) Los sistemas de ASR continua, que sirven para procesar frases más largas, tipo *IBM Via Voice* o *Dragon's Naturally Speaking*.

Como regla general, un reconocedor automático del habla no puede procesar correctamente el habla que difiere de la manera con la que el habla ha sido diseñada. Esta es una de las razones por las que los sistemas comerciales de dictado, cuando han sido pensados para inglés americano estándar, tienen una actuación pobre cuando encuentran muestras de habla acentuada tanto por hablantes no nativos como por hablantes de otros dialectos.

LaRocca (1999) también exploró las posibilidades del reconocimiento del habla e indicó que desde 1994 ha sido posible trabajar con esta tecnología con un coste menor. El uso de la ASR trajo nuevos beneficios al CALL en general y al autoaprendizaje en general. Los alumnos pueden practicar la lengua oral con una retroalimentación inmediata. De todas maneras, hace hincapié en trabajar, especialmente a nivel principiante, con reconocedores de voz híbridos, que procesan señales avanzadas juntamente con enlaces medidos de manera tradicional. Según este

autor, es de vital importancia el que cada lección tenga su propio vocabulario para producir reconocimiento de voz y esas producciones deben ser cortas. La forma en la que LaRocca utiliza el reconocimiento de voz es diversa:

- a) La forma más común es cambiar el teclado y el ratón por el micrófono con el paradigma de las respuestas múltiples. Según este paradigma, el programa hace escoger al alumno entre tres o cuatro alternativas por respuesta, entre las cuales el aprendiz ha de leerla en voz alta. El sistema está entrenado para reconocer esas alternativas.
- b) Los alumnos están obligados o incitados mediante texto, vídeo o audio a realizar el ejercicio.
- c) Para trabajar el análisis contrastivo de la pronunciación, se presentan dos palabras para trabajar los pares mínimos.

En relación al entrenamiento de la pronunciación, los tutores de pronunciación con voz interactiva hacen repetir a los alumnos palabras y frases. La clave para que ello sea un éxito es la retroalimentación correctiva, más específicamente un tipo de retroalimentación que no confie en la propia percepción del alumno. Mientras que la gramática y el vocabulario de un tutor de pronunciación son comparativamente simples, el procesamiento del habla tiende a ser complejo y se debe adaptar para reconocer y evaluar el habla del alumno, que no siempre es fluida. Un tutor de pronunciación, como contraste, debe entrenarse en reconocer y corregir las desviaciones de pronunciaciones nativas comunes.

Ehsani y Knodt (1998) citan a Berstein (1997), Franco *et al.* (1997), Kim *et al.* (1997) y Witt y Young (1997), que tienen diversas técnicas para reconocer de manera automática y anotar el habla no nativa. En

términos generales el procedimiento consiste en construir modelos de pronunciación nativa y entonces medir las respuestas no nativas con los modelos nativos. Ello requiere modelos entrenados tanto en respuestas con datos de habla nativas como no nativas en la lengua de estudio y suplementadas por una serie de algoritmos –un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema- para medir las variables acústicas que se han probado útiles en distinguir el habla nativa de la no nativa. Estas variables incluyen la latencia de respuesta, la duración segmental, las pausas entre palabras (en frases), la probabilidad espectral y la frecuencia fundamental (F0). En la retroalimentación suprasegmental, en la que la entonación y la acentuación son importantísimas, también la F0 es una de las correlaciones acústicas principales. En la figura 23, se ofrece una reproducción de su diseño de reconocimiento del habla, donde se puede observar su complejidad.

Existen tres aspectos del habla que reciben atención especial: el desarrollo de vocabulario, la práctica conversacional y la pronunciación, incluyendo la práctica de pares mínimos y la anotación de la pronunciación.

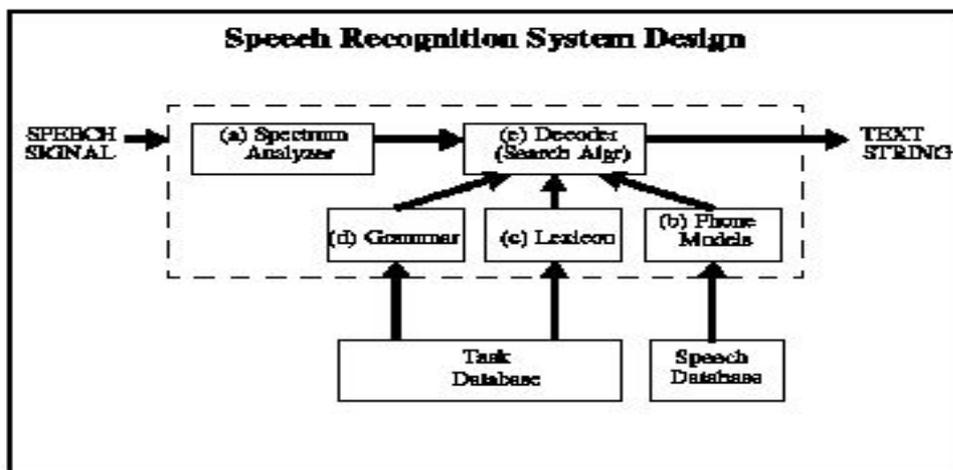


Figura 23. Ehsani y Knodt (1998:48). Diseño de un sistema de reconocimiento del habla

### **2.12.1 Los porcentajes de reconocimiento de voz**

Ainsworth y Greenberg (1996), en sus estudios sobre el reconocimiento automático del habla, dicen que trabajando de manera tradicional y con condiciones limitadas de vocabulario, suele dar un 98-99% de corrección. Ese porcentaje se reduce a un 88-95% cuando el material consiste en habla individual y teniendo siempre en cuenta que la grabación se efectúe en un ambiente acústico relativamente libre de ruido.

Sin embargo, bajo muchas condiciones que simulen la típica comunicación humana, la ASR funciona de manera mucho más pobre. El sistema informático de reconocimiento del habla suele fallar cuando se confronta con materiales hablados presentados bajo condiciones auditivas más realistas, en las cuales el retumbo y el alto nivel de interferencia acústica de fondo suelen ocurrir. Eso hace reducir el nivel de precisión en la palabra a un 20-50%.

Dos años más tarde, Greenberg *et al.* (1998), se refieren de una manera más técnica a la misma cuestión diciendo que cuatro canales espectrales distribuidos en un margen de habla de entre 0.3 a 6 khz son suficientes para que el ser humano pueda escuchar y decodificar materiales a nivel de frases con un 90% de precisión, aunque más de un 70% del espectro se pierde. El nivel de reconocimiento a nivel de palabras es relativamente alto (de un 60 a un 83%) cuando sólo se presentan dos o tres canales simultáneamente, a pesar de que la inteligibilidad de las mismas, presentadas de manera aislada, es de menos del 9%.

Bates (2003:1-2), en la introducción de su tesis sobre la variabilidad de pronunciación para modelos de reconocimiento de voz, en lo que se refiere a los datos de fiabilidad en reconocimiento de palabras con alta

calidad de sonido, comenta que el porcentaje alcanzado es mayor de un 90% en vocabularios de más de 50.000 palabras. Desafortunadamente, los resultados no son tan buenos cuando el habla es espontánea, donde se ha encontrado más de un 20% de errores al utilizar el teléfono y de más de un 35% con múltiples hablantes utilizando micrófonos ajustados a la cabeza.

### **2.12.2 Los reconocedores de voz de habla aislada y continua**

Faudez (2000:98) hace una distinción según la manera de enlazar las palabras en el habla:

- a) Reconocimiento del habla aislada: deben realizarse pausas suficientemente largas entre palabras (típicamente más de 200 ms), como si se realizara un dictado palabra por palabra. La ventaja es que resulta más sencillo detectar los inicios y finales de palabra que mediante habla continua, puesto que se han añadido silencios entre palabras. El inconveniente es que requiere la cooperación del locutor, que debe cambiar la manera normal de hablar.
- b) Reconocimiento del habla continua: es el sistema más complejo, ya que el reconocedor se verá afectado por efectos coarticulatorios. Un mismo fonema no siempre suena igual; depende de cuál haya sido el fonema anterior. La ventaja es que trabaja con la manera normal de hablar entre personas.

Según Ehsani y Knodt (1998), programas de reconocimiento de voz de manera continua como *Dragon's Naturally Speaking* o *IBM Via Voice*, por ejemplo, actúan con una precisión de reconocimiento de sólo un 60% a 80% (dependiendo del acento, ruido, tipo de producción, etc.).

### **2.12.3 Los reconocedores de voz con puntuación**

Llisterri (2003:14) trata los criterios de parecido o aceptación que se suelen aplicar en los programas educativos. En ellos, el reconocedor de voz ha de aceptar o rechazar enunciados con más o menos parecidos al original, y proporciona al alumno una puntuación, que no es otra cosa que la “distancia” entre el modelo que el usuario ha escuchado e intenta imitar y su propia producción. Dicha puntuación se establece numéricamente (con una escala de 1-10, de 1-5, etc.), pudiéndose en algunos casos (programas de *Auralog*), aumentar el criterio de evaluación en determinados programas para que el aprendiz intente realizar producciones mucho más precisas. Al referirse al proceso de reconocimiento en general y al específico, es decir, a los problemas localizados, he aquí una cita explicativa:

*“La comparación entre el enunciado almacenado y el original requiere también una deducción, por parte del sistema de reconocimiento, de los elementos que forman el enunciado, de modo que no únicamente se realiza una comparación global, sino que se lleva a cabo lo que se conoce como una “alineación”, es decir, una sincronización entre el texto correspondiente al enunciado y la señal sonora. Esto permite señalar al estudiante aquellos fragmentos de su producción en los que se produce una discrepancia más fuerte con respecto al modelo, de modo que puede concentrarse en ellos a la hora de repetir para mejorar la semejanza con el enunciado original”.*

Llisterri (2003:15)

Rypa y Price (1999) dicen al respecto que la anotación de la pronunciación en una actividad sólo se establece después de que se complete totalmente una lección y se hayan recogido suficientes datos con una puntuación aceptable. Para validar la puntuación automática de la pronunciación, en este sistema, se recoge tanto el habla nativa como la no nativa y se crea una base de datos de porcentajes realizada por expertos (por ejemplo, instructores lingüísticos).

### **2.12.4 La tecnología HMM (Hidden Markov Modelling)**

Varios autores (Ehsani y Knodt, 1998; LaRocca, 1999; Bernstein y Franco, 1996) han alabado la tecnología HMM (*Hidden Markov Modelling*) empleada en los reconocedores de voz. Dicha tecnología debe su nombre al matemático ruso Andrei A. Markov (1856-1922) y, según Rabiner (1989), tiene su origen en las publicaciones de una serie de documentos clásicos por Baum *et al.* (1966) y se implementó para aplicaciones de procesamientos del habla gracias a Baker (1975), en la Carnegie Mellon University (CMU) de Pittsburg, en Estados Unidos. Dicho modelo aplica estadísticas sofisticadas y computaciones probabilísticas al problema de enlazar modelos a nivel de fonemas y sílabas.

Para Ehsani y Knodt (1998), un reconocedor del habla HMM consta de cinco componentes básicos:

- a) Un analizador de señal acústica que computa una representación espectral de habla que se acaba de procesar.
- b) Un juego de modelos fónicos (HMMs) entrenados con grandes cantidades de datos de habla.
- c) Un léxico para convertir secuencias fónicas con fonemas y sílabas en palabras.
- d) Un modelo estadístico de lengua o enlace gramatical que defina la tarea de reconocimiento en términos de combinaciones de palabras legítimas a nivel de frase.
- e) Un decodificador, que es un algoritmo de búsqueda para computar el mejor enlace entre una producción hablada y su correspondiente palabra.

## 2.13 Los micrófonos

Por lo que se refiere a la calidad de la entrada de datos para una actuación óptima de reconocimiento, la señal de entrada del habla debe tener calidad acústica.

No sólo la interferencia de ruido puede afectar a la calidad del habla. Existen otros factores como el tipo de tarjeta de sonido, de micrófono empleado, o si el habla se efectúa mediante un pre amplificador. Algunas tarjetas de sonido tienen amplificadores internos, pero tienden a ampliar el sonido junto con el habla. La amplitud de la señal ha de ser ajustada cuidadosamente para su mejor reconocimiento.

Si observamos la figura 24, podemos ver los diversos mecanismos que influyen en un reconocimiento de la voz. Éste es un aspecto en el que el micrófono puede representar una tremenda diferencia en la actuación de reconocimiento. Muchos reconocedores actúan mejor cuando se emplea un micrófono ajustado a la cabeza con un cancelador de ruido. No sólo filtran el ruido exterior, sino que también, al estar colocados en la cabeza, se aseguran que la distancia entre la boca del hablante y el micrófono se mantiene constante y la amplitud es estable a través de las producciones.

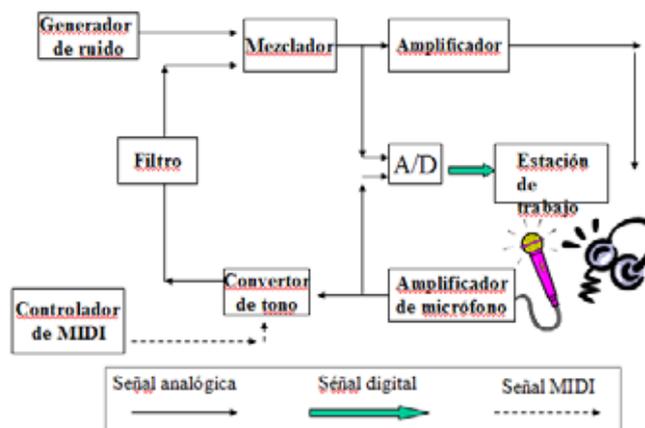


Figura 24. Kawahara (1996:15). Análisis de la sensibilidad de los parámetros del habla

Stibbard (1996) considera de crucial importancia la colocación del micrófono y dice que se ha de sostener el micrófono cerca de la boca, pero excesivamente, ya que algunos sonidos como los sibilantes o las vocales (altas – *loud*) producen una sobrecarga.

En este apartado se ha ofrecido una breve referencia teórica. Cuando tratemos las recomendaciones metodológicas, este aspecto se ampliará en el apartado 4.3.2

## **2.14 El análisis de errores**

Un sistema de reconocimiento de voz ha de ser capaz de reconocer los errores tanto segmentales como prosódicos. Al tratar la percepción del habla, Brett (1999) cita a Rost (1990), quien, al referirse a las características de la señal acústica, los fonemas y su articulación, comenta que la realización de fonemas ingleses está afectada por el contexto de los fonemas en los que aparecen, pudiéndose producir variaciones libres, asimilaciones, reducciones y elisiones. Las personas producen diferentes resultados acústicos dependiendo de su edad, sexo, región y características socioculturales, afectados también por la manera de expresarse (por ejemplo, placer, agitación, etc). No obstante, los fonemas son aún percibidos de manera exitosa. Igualmente, Cutler (1997), utilizando el *MARSEC Spoken Corpus*<sup>21</sup>, ha calculado que aproximadamente el 85% de las unidades polisilábicas en inglés contienen palabras incrustadas o encajadas, por ejemplo: ‘bracing’ contiene ‘brace’, ‘race’, ‘ace’, ‘racing’ y ‘sing’.

---

<sup>21</sup> MARSEC (Machine Readable Spoken English Corpus) está basado en muestras del habla inglesa en acento común del inglés británico del sur y ha sido desarrollado por Lancaster y IBM basándose en el SEC (Spoken English Corpus). La diferencia entre el primero y el segundo es que el MARSEC incorpora un segundo CD-Rom con grabaciones acústicas (el SEC carece de ellas) añadidas e incluye una alineación entre la palabra y el tiempo entre las transcripciones y la señal acústica.

Chanier *et al.* (1992:134) tratan el análisis de errores en los sistemas asistidos por ordenador. Este análisis, conocido en inglés como *parsing*, procesa el significado de los diversos comandos que utilizan algunas estrategias para reconocer una producción como errónea, localizar este error respecto a la gramática computacional y, a veces, poder remediarlo.

### **2.14.1 El alineamiento forzado de errores**

Un reconocedor de voz basado en un modo de alineamiento forzado (*forced alignment mode*) puede calcular los errores. El sistema enlaza el texto de la señal entrante como la señal usando información sobre el contenido lingüístico que ya ha sido almacenado en la memoria. Entonces, después de compararla con la nativa, se pueden identificar y localizar.

Varios autores tales como Horak (2001), Eskenazi y Hansma (1998) y Eskenazi (1999) se han referido a este tipo de tecnología. La siguiente cita resume su uso:

*"Given the sentence that the speaker was to say, the speech recognizer can return the scores of the words and the phones in the utterances. By comparing the speaker's recognition scores to the mean scores for native speakers for the same sentence pronounced in the same speaking style, errors can be identified and located. For prosody errors, only duration can be obtained from recognizer output. That is, when the recognizer returns the duration of the phones".*

Eskenazi y Hansma (1998:78)

### **2.14.2 Diversos estudios**

Sustarsic (2001) y Petek (1999) han elaborado estudios sobre áreas problemáticas en el reconocimiento de voz en inglés. Teniendo en cuenta algunos de los programas de reconocimiento de voz, tales como *Via Voice*

(patentado por IBM) o *Dragon's Naturally Speaking* (Dragon System Inc.), Sustarsic (2001) realizó diversas aproximaciones al análisis de los errores para trabajar pronunciación y encontró que la mayoría de errores de reconocimiento consistían en omisiones, adiciones y confusiones o malas interpretaciones. Este último punto se desarrolló en cuatro subapartados: vocabulario técnico, abreviaciones, vocabulario con nombres propios y vocabulario general.

Lo que es importante destacar es que, si el problema es el reconocimiento del habla de manera continuada -entiéndase aspectos suprasegmentales- la fiabilidad de los datos en los espectrogramas puede estar influenciada por los siguientes factores, que podemos resumir en la siguiente cita:

*“A word of caution though: speech recognition programs deal better with relatively slow, clear and “careful” (non-assimilated) style of speech, which might encourage students to train the “foreign language” word-by-word pronunciation – precisely the opposite of what we want them to achieve. On the other hand, this may not be such a serious problem if we instruct students to apply the more natural colloquial style already in the process of training the computer to their voices, following the simple philosophy that the computer is the tool of the user and not the other way round”.*

Sustarsic (2001:50)

Si bien el problema descrito es aquí mucho más preciso y no directamente relacionado con un programa de mejora de la pronunciación, debemos tener en cuenta que la producción por parte de los alumnos de frases de mayor o menor complejidad puede resultar difícil; debe realizarse un seguimiento del alumno y no permitir que se puedan producir fosilizaciones.