

# La puissance du multimédia au service de l'apprentissage des langues étrangères

## L'utilisation de la reconnaissance vocale

Nagi Sioufi (80)

### Des méthodes traditionnelles vers les méthodes multimédias

Depuis le début des années 90, l'apprentissage des langues étrangères subit de profondes transformations du fait de l'avènement du multimédia. L'apport des technologies nouvelles dans ce domaine rend les outils traditionnels de plus en plus obsolètes, et ceci bien plus que dans les autres domaines. Cela explique le succès commercial des CD-ROM d'apprentissage des langues qui représentent aujourd'hui plus de 30 % du marché des CD-ROM éducatifs.

Les outils traditionnels s'appuient soit sur un support papier (livre, dictionnaire, encyclopédie, etc.), soit sur un support audio (cassette, CD audio, etc.), soit sur un support vidéo (films en version originale, CNN, etc.).

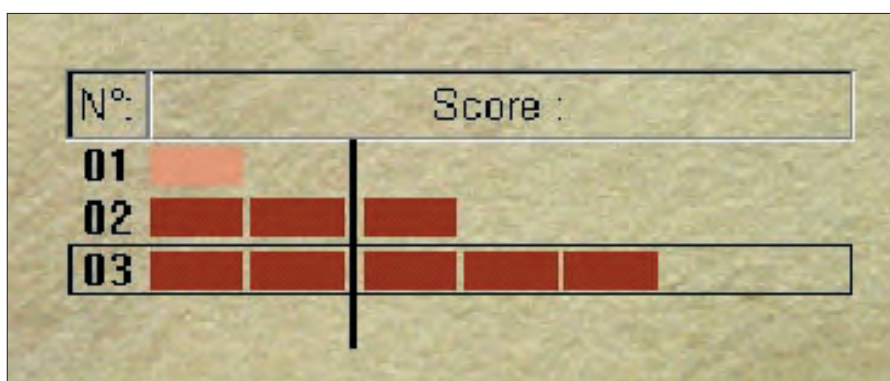
Ces différents supports restent répandus et utilisés là où l'ordinateur ne peut être présent. Mais la miniaturisation des composants et le développement des différentes formes d'ordinateurs itinérants raréfient ces situations.

L'exemple des laboratoires de langues à base de cassettes est significatif de cette évolution. Ces équipements nécessitaient des investissements élevés et continuaient à être installés jusqu'au milieu des années 90. Leur remplacement par des solutions à base d'ordinateurs en réseau, équipés des logiciels adéquats, est devenu irréversible. Ces laboratoires multimédias incluent toutes les fonctions de leurs prédécesseurs, en offrent des centaines d'autres supplémentaires, s'appuient sur des architectures standard et évolutives et coûtent moins cher du fait de la baisse des prix du matériel informatique.

Les cours avec un professeur particulier sont souvent très efficaces mais deviennent très rapidement excessivement coûteux. Les tarifs peuvent baisser si le nombre de participants s'élève, mais il est prouvé que l'efficacité est inversement proportionnelle à ce nombre. De plus, les cours imposent une contrainte géographique et temporelle : l'apprenant doit se déplacer dans un lieu qui ne l'arrange pas toujours, et ceci à une heure déterminée et selon une fréquence qui lui conviennent rarement lorsqu'il exerce une activité professionnelle.

Le multimédia fusionne tous les supports existants et ajoute l'interactivité. Son incidence sur l'apprentissage devient considérablement plus élevée de ce fait. La transition des méthodes de langues vers le multimédia a cependant donné lieu à trois générations successives.

Lors de la **première génération**, les premiers balbutiements sont apparus sous forme de transfert de contenus pédagogiques d'un ou de plusieurs supports traditionnels vers le support multimédia. Par exemple, au cours du transfert d'une méthode sur cassettes et du livret d'accompagnement vers le CD-ROM, un logiciel de navigation plus ou moins performant pouvait donner l'impression d'une exploration facile du contenu. Certains exercices pédagogiques transférés du support papier, comme le texte comportant des "trous" ou l'exercice d'association de synonymes, apportaient une première interactivité avec la correction automatique par l'ordinateur de l'activité de l'apprenant. Le transfert de la fonction "enregistrez-vous" des méthodes sur cassettes vers l'ordinateur rendait la manipulation bien plus commode sans pour autant apporter une grande valeur



ajoutée pédagogique. De même, les méthodes d'Enseignement assisté par ordinateur (EAO) à base de vidéodisques se sont diffusées à cette période. Elles permettaient essentiellement de visionner un film en version originale et de bénéficier de tout un ensemble de fonc-

tions facilitant la compréhension de l'apprenant : avant, arrière, stop, pause, avance rapide, traduction, sous-titres, explications, etc.

La **deuxième génération** de produits a dépassé ce stade du transfert de support et a commencé à utiliser toute

la puissance des nouvelles technologies. L'utilisation de la reconnaissance vocale pour aider l'apprenant dans l'apprentissage de l'expression orale et de la prononciation a constitué une véritable révolution dans la mesure où l'outil obtenu permettait d'apporter une solution efficace à un problème mal résolu jusque-là. La technologie de la reconnaissance vocale qui est détaillée dans la suite de ce document est d'ailleurs vite devenue la technologie clé dans l'apprentissage des langues. De la même façon, d'autres produits ont exploité la puissance de l'ordinateur pour apporter une solution à un besoin pédagogique spécifique. Les outils intégrant des correcteurs orthographiques et grammaticaux ont aidé les moins novices à mieux rédiger en repérant les fautes les plus évidentes. Les logiciels de suivi ont permis de garder une trace du comportement de l'apprenant, de l'évaluer et de définir progressivement des enseignements individualisés adaptés aux besoins de chacun.

La **troisième génération** est la plus récente. Elle comporte des méthodes complètes qui couvrent tout le processus d'apprentissage d'une langue étrangère : expression orale, expression écrite, compréhension orale, compréhension écrite, grammaire, vocabulaire. Elles s'adaptent à tous les niveaux, du débutant à l'élève confirmé, et exploitent toutes les dernières avancées technologiques : reconnaissance vocale, vidéo MPEG, accès Internet, diagnostic intelligent, etc.

## Évolution de l'apprentissage des langues étrangères à l'oral

L'apprentissage des langues étrangères à l'oral a connu différents stades de développement majeurs.

Au tout début, l'apprenant avait un seul moyen à sa disposition pour se perfectionner à l'oral : reproduire le plus fidèlement possible la **voix de son professeur**.

Puis il a eu la possibilité d'**enregistrer** sa propre voix et de se réécouter. D'analogique l'enregistrement

devint numérique, mais malgré une amélioration qualitative, cette technique restait limitée par la perception auditive du sujet.

Au début des années 90, la société Auralog a introduit pour la première fois la **technologie de la reconnaissance vocale** dans ses méthodes, permettant ainsi à l'apprenant de dialoguer librement avec l'ordinateur, sans manipulation du clavier ni de la souris, et d'obtenir une évaluation automatique de sa prononciation.

## Définition et classification de la reconnaissance vocale

La reconnaissance vocale désigne l'ensemble des technologies permettant à une machine de reconnaître la parole. Les technologies de reconnaissance vocale modernes sont basées sur une analyse des phonèmes de la phrase prononcée.

Trois critères permettent de distinguer les moteurs de reconnaissance vocale.

1. Le premier est le **type de discours** reconnu (mot à mot ou parole continue) par le logiciel :

- soit les mots sont prononcés **isolément**, c'est-à-dire séparés par des périodes de silence,
- soit le discours se fait en **continu** et permet ainsi une prononciation naturelle.

2. Le deuxième critère distingue les moteurs de reconnaissance vocale fonctionnant uniquement sur les **mots**, de ceux fonctionnant sur les **phrases complètes**.

3. Le troisième critère est le degré de dépendance vis-à-vis du locuteur :  
 - système monolocuteur : système adapté ou adaptable grâce à un processus d'entraînement à un locuteur particulier ("**speaker dependent**"),  
 - système multilocuteurs ("**speaker independent**") : reconnaissance vocale fonctionnant pour tout un groupe de locuteurs, ce groupe pouvant, à l'extrême, intégrer n'importe quelle personne parlant la langue apprise (indépendamment de son âge, de son sexe, du timbre de sa voix...).



Les meilleurs CD-ROM du marché utilisent les moteurs de reconnaissance vocale les plus sophistiqués : l'apprenant peut perfectionner sa prononciation sur des mots ou des phrases complètes, au rythme qu'il désire. De plus, les logiciels de type "speaker independent" sont immédiatement utilisables par tous les locuteurs, sans entraînement préalable.



## La reconnaissance vocale : vingt ans d'évolution technologique

La technologie de la reconnaissance vocale a évolué, passant du hardware (cartes) au software. Aujourd'hui, l'apprenant prononce librement une phrase et le logiciel note la qualité de la prononciation par une analyse complexe des phonèmes.

Voici la chronologie des différentes étapes du développement de la reconnaissance vocale sur les vingt dernières années.

### Début des années 1980

Les premiers systèmes **monolocuteurs** de reconnaissance vocale sont appliqués sur des machines industrielles.

### 1985

C'est environ à cette date qu'apparaissent les premières **cartes** équipées de DSP (Digital Signal Processor), permettant d'implémenter une technologie de reconnaissance vocale sur un PC. Ces technologies sont très complexes à mettre en œuvre (test sur un panel significatif de locuteurs). Ces cartes ne sont pas diffusées au grand public.

### 1991

La première application d'apprentissage de langues fondée sur la reconnaissance vocale est créée.

### 1994

C'est l'arrivée des cartes DSP **grand public** (sous Windows™).

### 1995

Des **solutions logicielles** issues de la technologie des cartes font leur apparition. Le premier logiciel grand public sous Windows™ est lancé avec un système de reconnaissance vocale multilocuteurs fonctionnant sur des phrases complètes. La reconnaissance vocale permet au logiciel d'analyser la prononciation de l'apprenant et de l'évaluer grâce à un système d'évaluation sophistiqué.

### 1996

La reconnaissance vocale est exploitée différemment selon son utilisation : méthode de langues ou logiciel de dictée vocale.

*Parallèlement au marché des langues, les premiers logiciels de dictée vocale sont lancés sur le marché. Le principal éditeur est alors DragonSystems. La reconnaissance vocale se fait mot à mot.*  
**1998**

Des méthodes complètes de langues sont lancées, dans lesquelles, pour la première fois, la reconnaissance vocale est appliquée aux exercices linguistiques.

*Sur le marché de la dictée vocale apparaissent les premiers logiciels autorisant la dictée **en continu** (le principal logiciel est IBM ViaVoice).*  
**1999**

Une innovation technologique révolutionnaire, **SETS (Spoken Error Tracking System)** apparaît et permet de localiser les défauts de prononciation.

La reconnaissance vocale fonctionnait déjà de façon continue sur des phrases complètes et indépendamment du locuteur. Restait encore un problème qu'aucune équipe de développement n'avait résolu jusqu'à présent : identifier l'élément mal prononcé dans une phrase complète.

Après plusieurs années de recherche, des ingénieurs ont mis au point une technologie qui permet de détecter automatiquement l'erreur de prononciation dans une phrase. Cette technologie exclusive, baptisée **SETS (Spoken Error Tracking System)**, fait l'objet d'un dépôt de brevet. En repérant les points faibles dans la pro-

nonciation de l'apprenant, **SETS** représente une avancée décisive dans l'apprentissage d'une langue étrangère sur ordinateur.

## Exploitation de la reconnaissance vocale dans les méthodes de langues

### 1. Le dialogue interactif

Grâce à la technologie avancée de la reconnaissance vocale, l'utilisateur engage un véritable dialogue avec son PC. Les réponses de l'utilisateur orientent la conversation.

Suivant son niveau, l'apprenant paramètre la reconnaissance vocale pour la rendre plus tolérante ou plus exigeante quant à la qualité de sa prononciation.

### 2. La prononciation de la phrase ou du mot

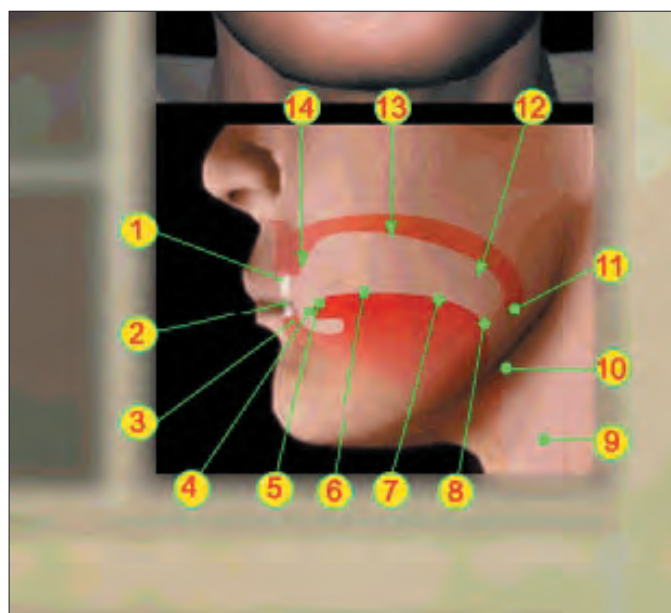
L'utilisateur s'entraîne à prononcer une phrase ou un mot et obtient une note (score) lui permettant d'évaluer la qualité de son accent, de sa prononciation et de son intonation.

### 3. L'affichage de la phrase ou du mot prononcé

Le logiciel offre – en plus de l'évaluation – la possibilité de visualiser

précisément sa prononciation et son *intonation*. Deux types de représentations (un graphe et une courbe) sont disponibles. L'utilisateur peut les afficher simultanément ou séparément.

Le graphe vocal indique l'amplitude de la voix en fonction du temps (notion d'énergie). Il repré-



sente l'intensité sonore de la voix et donne des indications sur la structure de la prononciation.

La courbe de la fréquence fondamentale reproduit les variations de la fréquence de la voix en fonction du temps (notion de hauteur). Cette courbe, complémentaire du graphe vocal, permet à l'apprenant de comparer précisément son intonation à celle du modèle (aigu/grave).

Les meilleurs logiciels proposent des fonctions qui **évaluent** et permettent de visualiser la **prononciation** et **l'intonation** aussi bien pour des **phrases complètes** que pour des mots.

#### 4. Les animations de phonèmes en 3D

Grâce à des animations phonétiques en images de synthèse, l'apprenant peut visualiser avec une grande précision les mouvements articulaires pour mieux les reproduire.

Cette technologie d'exception aide à comprendre le fonctionnement articulaire de la production de certains sons et à les mémoriser, afin de mieux les reproduire.

#### Des exemples d'application

Le nouvel aéroport de Hong-Kong, Chek Lap Kok, utilise des CD-ROM *Skytalk* avec reconnaissance vocale pour son centre de formation au contrôle du trafic aérien. Ces CD-ROM permettent aux pilotes et aux aiguilleurs du ciel de perfectionner leur expression orale en anglais. *Skytalk* invite l'utilisateur à jouer le rôle d'un pilote ou d'un aiguilleur du ciel en prenant part à un dialogue interactif, inspiré d'un réel échange radio. Les conversations, basées sur des échanges et des situations très caractéristiques, permettent à l'apprenant de se familiariser avec le vocabulaire propre à l'aviation. Grâce à la reconnaissance vocale, qui favorise un apprentissage personnalisé, l'ordinateur joue le rôle de professeur particulier : il pose des questions à l'utilisateur, réagit en fonction de ses réponses et évalue sa prononciation au fil de la conversation. Dans l'exercice de prononciation, celui-ci s'entraîne à prononcer les expressions

difficiles et surmonte ainsi rapidement ses hésitations. Très interactive, cette méthode enseigne aux apprenants à comprendre et à réagir rapidement dans des situations très diverses, facteur primordial pour garantir la sécurité des communications sol/air. Pour bon nombre d'organismes, la communication en anglais entre les pilotes et les aiguilleurs du ciel est devenue une préoccupation majeure. L'utilisation de ces outils s'inscrit donc, pour ces centres de formation, dans un effort continu d'amélioration de la sécurité aérienne. D'autres organismes utilisent la même méthode, tels que la Singapore Aviation Academy et l'organisme de formation italien, la Ente Nazionale di Assistenza al Volo.

En France, le ministère de l'Éducation nationale a équipé la totalité des Instituts universitaires de formation des maîtres (IUFM) et des Centres régionaux de documentation pédagogique (CRDP) en CD-ROM *Tell me More*. De la même façon, en Espagne, la Junta de Andalucia a acquis plus de 500 licences afin de couvrir l'ensemble du système scolaire secondaire d'Andalousie.

De nombreuses entreprises s'intéressent également à ces méthodes, qui constituent un outil de formation performant pour leurs salariés : citons Mercedes-Benz, qui équipe la totalité de son groupe avec plus de 2 000 licences ou l'Union des Banques suisses, avec plus de 500 licences.

Dernier équipement en date, la compagnie aérienne nationale Air France vient d'acquérir 1 000 licences de CD-ROM : 900 licences "classiques" seront réparties dans les centres de formation ou dans les antennes d'Air France du monde entier, et 100 licences "déportées" permettront aux employés de la compagnie d'équiper leurs propres micro-ordinateurs. L'utilisation de licences déportées permettra donc aux apprenants de poursuivre leur apprentissage à domicile ou lors de fréquents déplacements. Cet équipement porte sur les 6 langues d'apprentissage principales (anglais britannique, anglais américain, espagnol, allemand, italien, français langue étrangère), selon tous les niveaux (débutant, moyen, confirmé, affaires).

#### Glossaire

- **Speech recognition – Reconnaissance vocale**

Ensemble des technologies permettant à une machine de reconnaître la parole.

- **Difficulty level of the speech recognition – Niveau de difficulté de la reconnaissance vocale**

Niveau à partir duquel l'ordinateur accepte un mot ou une phrase comme suffisamment bien prononcé. Les produits Auralog permettent à l'utilisateur de moduler ce niveau, afin d'adapter la difficulté de prononciation.

- **SETS (Spoken Error Tracking System) – Technologie SETS**

Technologie exclusive d'Auralog permettant de détecter automatiquement l'erreur de prononciation dans une phrase complète.

- **Waveform – Graphe vocal**

Représentation graphique de l'amplitude de la voix en fonction du temps. Ce graphe donne une indication sur la structure de la prononciation.

- **Pitch curve – Courbe de la fréquence fondamentale**

Représentation graphique permettant de visualiser les variations de la fréquence de la voix en fonction du temps et ainsi d'améliorer l'intonation.

- **Word-by-word speech recognition – Reconnaissance vocale mot à mot**

Pour être reconnu par l'ordinateur, les mots doivent être prononcés isolément, c'est-à-dire entrecoupés de silences.

- **Continuous speech recognition – Reconnaissance vocale en continu**

Caractérise les moteurs de reconnaissance vocale les plus évolués. Le discours avec l'ordinateur peut se faire en continu.

- **Speaker-dependent system – Système monoclocuteur**

Système de reconnaissance vocale adapté ou adaptable via un processus d'entraînement à un locuteur particulier.

- **Speaker-independent system – Système multilocuteurs**

Reconnaissance vocale fonctionnant pour tout un groupe de locuteurs, ce groupe pouvant, à l'extrême, intégrer n'importe quelle personne parlant la langue apprise.

- **DSP (Digital Signal Processor) – Processeur de signal**

Processeur permettant d'implémenter une technologie de reconnaissance vocale dans un micro-ordinateur. Il a été remplacé par des solutions logicielles de reconnaissance vocale.