Société
de développement
des entreprises
culturelles
Québec

L'ingénierie linguistique et ses applications éventuelles à l'industrie du doublage

Rapport de recherche

Recherche et analyse
Patricia Gariépy, consultante

Direction des travaux et révision
Anne-Marie Gill, chargée de projets
Direction générale politique, communications
et relations internationales

Mai 2004

TABLE DES MATIÈRES

Résumé de la recherche

L'ingénierie linguistique est la science qui préside à toutes les applications informatiques relatives aux langues et à la parole. Trois secteurs de cette discipline scientifique concourent au développement de logiciels de traitement de la voix et de la parole : celui de la reconnaissance de la parole, celui de la synthèse de la parole et celui de la conversion des voix.

Malgré des avancées importantes de la recherche et la création de nombreux logiciels du traitement de la voix, l'ingénierie linguistique est confrontée à des problématiques importantes quand il s'agit de reproduire la prosodie du discours. La prosodie constitue en quelque sorte la musique d'une langue et se traduit d'une langue à l'autre, voire d'un accent à un autre, par une palette de sons différents (phonèmes) et un rythme qui varie en fonction de l'intonation, de l'accentuation et du débit (variation de la hauteur, de la durée et de l'intensité du son, des pauses, etc.) de l'interlocuteur d'une langue donnée. Par ailleurs, si la prosodie transmet de nombreuses caractéristiques socioculturelles d'un interlocuteur, elle traduit tout autant son état émotionnel et varie de façon subtile en fonction de ce dernier. La prosodie fait partie intégrante du « message » et un logiciel performant de production de voix artificielles devrait pouvoir la décoder et la reproduire au même titre que les autres composantes de la parole. Or, le nombre exponentiel de variables introduites par la prosodie dans la parole est telle que, jusqu'à maintenant, l'ingénierie linguistique n'arrive pas encore à établir, sauf pour de brefs segments, des modélisations qui puissent décoder et reproduire aisément — et dans toutes ses nuances —, cette composante du discours. La création de voix artificielles pour remplacer la voix humaine dans des productions audiovisuelles, notamment dans les œuvres de fiction, n'est pas encore possible, à tout le moins, à ce stade-ci des recherches.

Du côté des technologies de la conversion de la voix et de leur usage dans le doublage des productions audiovisuelles, la technologie n'est pas encore au point et, advenant qu'elle devienne performante, sa généralisation est peu probable compte tenu de la complexité des opérations et des coûts qu'elle sous-tend. Elle pourrait être utile toutefois dans la création de voix particulières.

La numérisation des activités humaines a connu des développements majeurs au cours des dix dernières années, et cette situation pourrait se poursuivre dans l'avenir. L'expérience et une certaine prudence avisée nous invitent à penser que tout demeure théoriquement possible. Aussi, est-il important de poursuivre un monitoring assidu des avancées des technologies de la voix pour anticiper, le cas échéant, ses applications éventuelles dans le domaine de la culture et des productions audiovisuelles.

Introduction

La numérisation des processus de production et de diffusion dans le domaine culturel a connu des développements majeurs au cours des dix dernières années, et il est à prévoir que cette tendance se poursuivra durant plusieurs années encore. En outre, certains intervenants prétendent que le développement de l'ingénierie linguistique pourrait ouvrir la porte à l'utilisation des voix artificielles dans le domaine du cinéma et de la production télévisuelle, notamment dans les activités de doublage des productions. Préoccupée par la question, la Commission du doublage, en concertation avec le Conseil national du cinéma et de la production télévisuelle (CNCT), ont souhaité disposer de renseignements supplémentaires sur le développement des technologies dans ce domaine.

La recherche a donc pour but d'établir un portrait du développement des technologies de la conversion de la voix et de vérifier leur efficience réelle et possible dans les activités de doublage. Elle vise plus spécifiquement les objectifs suivants :

- disposer d'un portrait du développement de l'ingénierie linguistique et des nouvelles technologies de conversion de la voix dans le secteur du doublage (développement de logiciels), décrire les processus qu'elles sous-tendent, qualifier leur efficience à ce jour ainsi que dans un proche avenir;
- déceler les impacts que pourraient avoir ces nouvelles technologies sur :
    1) le processus et les coûts de production du doublage;
    2) les entreprises de doublage;
    3) le travail — droits et rémunération — des comédiens (comédiens de l'œuvre originale et comédiens doubleurs);
    4) les producteurs et les distributeurs des œuvres.
- connaître l'intérêt et les intentions des *majors* à l'égard de ces développements technologiques ou de tout autre élément qui pourrait y être relié;
- dégager les principaux constats de la recherche et proposer le mode de suivi nécessaire au dossier.

Le contenu du rapport

Le chapitre 1 porte sur le processus de production d'un doublage d'une œuvre de fiction afin de conférer à tous les lecteurs une même information de base qui leur permettra d'apprécier à sa juste valeur les changements qui pourraient ou non survenir dans les activités de doublage. Les chapitres suivants (chapitres 2 à 5) passent en revue trois secteurs de l'ingénierie linguistique, soit : la reconnaissance vocale, la synthèse vocale et la conversion des voix. Bien que distincts, nous avons cru nécessaire de bien camper les caractéristiques de ces trois champs d'étude parce qu'ils sont reliés entre eux et que l'évolution des recherches dans l'un d'eux peut aisément avoir des répercussions dans celle d'un autre. Enfin, les derniers chapitres (6 à 8) présentent les impacts éventuels que pourraient avoir des technologies performantes de conversion des voix sur le doublage, dégagent les conclusions générales de la recherche et proposent des modes de suivi de ce dossier.

La méthodologie

Pour compléter notre recherche, nous avons d'abord consulté, puis analysé la littérature scientifique et commerciale disponible dans Internet et nous avons interviewé quelques spécialistes en ingénierie linguistique au Québec, ce qui nous a permis de faire le tour de la question et d'établir les principaux enjeux pour l'industrie du doublage. Les limites de la recherche sont attribuables au développement parfois rapide des découvertes dans le domaine des technologies et à la tenue confidentielle de certaines informations jugées stratégiques par des entreprises pour la mise en marché éventuelle de nouvelles applications informatiques.

1. Le Doublage

1.1 Définition

Le doublage d'une œuvre de fiction est son adaptation dans une langue autre que la langue originale. On dit adaptation plutôt que traduction, car les dialogues doivent être adaptés aux mouvements des lèvres et au langage du corps des personnages à l'écran. Ils doivent également l'être au contexte socioculturel de la langue du doublage (par exemple dans le cas de blagues ou de jurons). Cette forme de doublage représente la plus grande part des activités de l'industrie et s'applique à toutes les productions de fiction que ce soit dans le domaine du cinéma ou dans celui de la télévision. C'est cette forme de doublage que nous décrivons au point 1.2 du rapport.

Il existe une autre forme de doublage, celle du documentaire ou du reportage. Comme il n'est pas nécessaire de synchroniser les dialogues avec les mouvements des lèvres, ce type de doublage est beaucoup plus simple et rapide que celui des œuvres de fiction.

1.2 Le processus et les étapes de production d'un doublage

Au Québec, comme en France, le doublage d'une œuvre de fiction est basé sur l'utilisation de la bande rythmo. La bande rythmo[1] est un ruban 35 mm clair sur lequel le texte de la version doublée est écrit à l'encre de chine. Lors de la séance d'enregistrement en studio, cette bande est projetée en synchronisme avec les images du film. Cette bande sert alors de point de repère au comédien qui peut alors lire et interpréter le texte en phase précise avec le déroulement du film.

Actuellement, le processus de production d'un doublage se décline en différentes étapes dont :

1.2.1 La détection

À partir du relevé des dialogues et d'une copie du film, le détecteur a pour mission de repérer le mouvement des lèvres des personnages à l'écran et de les noter au crayon à mine sur une bande qui fonctionne en synchro avec l'image. Cette bande (de l'amorce 35 mm blanche) se déroule à 1/8^{e} de la vitesse de l'image. Le texte abrégé, joint à des signes de synchro appropriés (ouverture/fermeture de bouche, début/fin de phrases, inspiration/expiration, rire, etc.) est inscrit sur la bande.

[1] Une modification du matériel utilisé dans la fabrication du ruban 35 mm clair entraînera d'ici peu des changements dans la **préparation** de la bande rythmo. Les impacts que ces changements auront sur certaines étapes du doublage peuvent être considérés comme relevant de l'ingénierie linguistique. Un exposé des changements affectant la bande rythmo est inclus à l'**annexe A**.

1.2.2 L'adaptation

À cette étape, un spécialiste, l'adaptateur, traduit et adapte les dialogues dans la langue du doublage en respectant les signes de synchronisme. Il transcrit ces nouveaux dialogues au crayon à mine sur l'amorce blanche, en haut ou en bas du texte préalablement détecté, en respectant les début et fin de phrase ainsi que les notes de synchronisme indiquées par le détecteur.

1.2.3 La calligraphie et la dactylographie

Une fois la nouvelle version adaptée, le texte est calligraphié au propre, à l'encre de chine, sur un ruban 35 mm clair qui servira à la projection en studio. Ce ruban constitue ce qu'on appelle communément la bande rythmo. Parallèlement, le texte est dactylographié selon les normes de calcul du paiement des comédiens (50 frappes à la ligne[2]). Ce texte dactylographié sert également à la préparation des sessions d'enregistrement.

1.2.4 Les sessions d'enregistrement

La préparation des sessions d'enregistrement se fait en organisant les convocations des comédiens de façon à favoriser l'interprétation des personnages ainsi qu'une utilisation efficace du temps des comédiens et des studios.

Par exemple, dans le but de faciliter le jeu des comédiens, il est préférable de les avoir ensemble en studio lorsqu'ils jouent des scènes où il y a interaction entre leurs rôles (disputes, batailles, scènes d'amour, etc.). Quant à la gestion du temps des comédiens, il faut tenir compte des règles établies aux ententes collectives en vigueur. D'une part, la structure de calcul du cachet des comédiens en doublage[3] est basée sur les plus élevés de :

- la durée de la convocation ou
- la durée effective des séances d'enregistrement ou
- le nombre de lignes.

D'autre part, il y a une limite au nombre de rôles qu'un comédien peut cumuler dans une même production[4].

Les comédiens, présents au même moment en studio, sont généralement enregistrés ensemble sur la même piste de son, à moins qu'une raison technique exige qu'ils le soient séparément. À la fin de l'enregistrement, il y a plusieurs pistes de son pour chaque scène du doublage.

---

[2] Article 1.22 de l'*Entente collective entre l'Union des artistes et l'Association des doubleurs professionnels du Québec*, du 1er mars 2003 au 28 février 2006
[3] Article 7.1.1 de l'*Entente collective*
[4] Article 5.1.5 de l'*Entente collective*

1.2.5 Le recalage

L'utilisation de la bande rythmo permet d'obtenir une bonne synchronisation des dialogues, à un taux qu'on pourrait évaluer entre 70 % et 80 %. Le recalage est l'étape par laquelle le synchronisme des dialogues et des images est rendu plus précis par un monteur sonore, appelé « recaleur ».

1.2.6 Le mixage

Le mixage est l'étape où toutes les pistes de dialogues sont calibrées et mixées à la bande-son internationale, laquelle contient la musique et les effets sonores du film. Cette bande-son, communément appelée « M & E », pour *musique et effets*, est fournie par le client et permet de donner l'illusion que les dialogues ont été dits dans les lieux que nous voyons à l'écran. Pour donner cette illusion, il faut utiliser des équipements périphériques, qui transforment la sonorité des dialogues enregistrés en studio, créant l'illusion de dialogues dits dans un lieu particulier. Par exemple, les personnages à l'écran se trouvent dans une grande caverne où il devrait y avoir beaucoup d'écho. Les dialogues de doublage enregistrés en studio ne comportent aucun écho. Il faut rajouter celui-ci au moyen d'équipements périphériques. Le mixage permet que les dialogues de doublage reproduisent l'écho qui convient à la caverne et que l'illusion soit complète.

Ainsi se termine le travail du doublage. Le mixage final du doublage est par la suite transféré sur la copie maîtresse ou originale de la production.

2. L'ingénierie linguistique

2.1 Définition et problématiques

L'ingénierie linguistique est l'application de la connaissance des langues et de la parole à l'élaboration de systèmes informatiques capables de reconnaître, de comprendre, d'interpréter et de produire du langage humain sous toutes ses formes : écrit, parlé ou codé autrement. C'est un domaine de recherche très complexe qui requiert des équipes pluridisciplinaires composées à la fois de linguistes, de documentalistes, de lexicographes, de traducteurs, de psychologues, d'informaticiens, de statisticiens, d'ingénieurs, etc.

Un premier niveau de complexité de cette science repose sur celle des langues. Système d'expression et de communication commun à un groupe social et à son histoire, chaque langue repose sur une organisation élaborée de mots, de phrases, de sons (phonèmes) et de sens différents, et présente des variations importantes sous ses formes parlée ou écrite. Les différences d'une langue à l'autre sont à ce point importantes qu'on ne saurait faire de traduction mécanique, au mot à mot, sans perturber, modifier, voire infirmer un propos.

Au plan informatique, on cherche d'ailleurs — et depuis longtemps — à mettre au point un système phonétique international mais on n'y est pas encore parvenu. Les systèmes informatiques doivent donc, jusqu'à maintenant, être élaborés sur les spécificités de chaque langue. Lorsque l'application vise le transfert d'une langue à l'autre, le code de conversion sera un système aussi complexe que celui qui aura été élaboré pour chacune des langues prises une à une. Notons que les systèmes de traduction gagnent en performance avec les années, mais l'intervention humaine demeure toujours nécessaire pour assurer la qualité et la validité de la traduction.

Le deuxième niveau de complexité auquel est confronté l'ingénierie linguistique concerne la langue parlée, la « parole ». Ici, le décodage ou la reproduction informatique nécessite non seulement une bonne connaissance de la langue et de ses composantes, mais aussi celle de son expression physique, la parole, notamment celle de la prosodie.

La prosodie[5] constitue en quelque sorte *la musique d'une langue* et se traduit d'une langue à l'autre, voire d'un accent à un autre, par une palette de sons différents (phonèmes) et un rythme qui varie en fonction de l'intonation, de l'accentuation et du débit (variation de la hauteur, de la durée et de l'intensité du son, des pauses, etc.) de l'interlocuteur d'une langue donnée. Par ailleurs, si la prosodie transmet de nombreuses caractéristiques socioculturelles d'un interlocuteur, elle traduit tout autant son état émotionnel et varie de façon subtile en fonction de ce dernier. La prosodie fait partie intégrante du « message », et un logiciel performant de production de voix artificielles devrait pouvoir la décoder et la reproduire au même titre que les autres composantes de la parole. Or, le nombre exponentiel de variables introduites par la prosodie dans la parole est telle que, jusqu'à maintenant, l'ingénierie linguistique n'arrive pas encore à établir, sauf pour de brefs segments, des modélisations qui puissent décoder et reproduire aisément et dans toutes ses nuances cette composante du discours. Nous verrons dans les chapitres suivants où en sont les recherches et les applications pratiques.

Le troisième niveau de complexité relève de la place de cette discipline scientifique au confluent des frontières de plusieurs autres. Ainsi, les modélisations établies dans le domaine de la linguistique, de l'électro-acoustique, de l'informatique, de la statistique, de la géométrie, de l'intelligence artificielle, de la psychologie, etc. procèdent par des angles et des méthodes d'analyse parfois très différents, et il est difficile de les concilier dans un même outil de synthèse. Les recherches et les collaborations demeurent néanmoins intensives et la production de logiciels de plus en plus performants se poursuit. Il est donc important de suivre les développements sur une base continue pour connaître les applications qui pourraient voir le jour, entre autres, dans le domaine de la culture.

[5] DUDLEY, John G. ; et Jocelyne DELAGE. « Ensemble des éléments phoniques (intonation affective, particularismes régionaux, accent tonique, montée mélodique, etc.) qui caractérisent le langage parlé ». *Le langage en suspens*, Saint-Lambert, Héritage, 1990. 232 p., (Neurolinguistique), page 211).

2.2 Les champs d'application et les secteurs de recherche

Les champs d'application de l'ingénierie linguistique sont nombreux. Mentionnons entre autres :

- l'interaction homme/machine tels les systèmes de téléphonie automatique (tel le service 411 de Bell Canada) et autres centres d'appels de service automatisés, ou encore les téléphones cellulaires avec reconnaissance vocale du carnet d'adresses;
- l'aide informatique à la traduction écrite (dictionnaires électroniques, « traducticiels » intégrés, etc.) ou parlée (traduction simultanée, sous-titrage simultané[6] pour mal-entendants, etc.);
- les outils d'aide à la rédaction (grammaire, dictionnaires, traitement de texte, etc.) et à la gestion d'information;
- l'aide à l'enseignement et au perfectionnement personnel;
- l'aide aux gens souffrant d'un handicap visuel, auditif, vocal au autre (exemple : un individu atteint d'un handicap vocal pourrait s'exprimer à l'aide d'un synthétiseur);
- les outils dans les domaines de la culture, des arts et du loisir, tels les systèmes de correction de la voix dans des enregistrements sonores ou de production de certaines voix automatisées dans les jeux vidéos;
- etc.

Comme on le constate à la lecture de cette liste, certains des champs d'application de l'ingénierie linguistique s'appliquent au langage écrit tandis que d'autres visent le langage parlé. Dans le cadre de notre rapport, nous nous sommes plus particulièrement intéressés aux champs d'application relatifs au langage parlé lesquels interpellent trois grands secteurs de recherche, soit ceux de :

- la reconnaissance vocale;
- la synthèse vocale;
- la conversion des voix.

3. La reconnaissance vocale

La reconnaissance vocale est « l'ensemble des techniques ayant pour but de permettre à un ordinateur de reconnaître les signaux émis par la voix humaine en vue d'en faire le traitement »[7].

Les premières tentatives d'utiliser la reconnaissance vocale ont été effectuées dans les années 40 par le Département de la défense aux États-Unis. Depuis ce temps, nous avons incorporé dans notre vie quotidienne beaucoup d'applications, par exemple :

- les systèmes de téléphonie;
- les téléphones cellulaires à reconnaissance vocale;
- les systèmes de dictée vocale associée à un traitement de texte;
- la possibilité de contrôler certains appareils à distance.

[6] TVA offre un tel service de traduction pour les informations météo en direct.
[7] *Le Grand Dictionnaire de terminologie* de l'Office de la langue française, www.granddictionnaire.com

3.1 La reconnaissance de la parole (ce qui est dit)

Les logiciels de reconnaissance vocale (ce qui est dit) fonctionnent de la façon suivante :

> « La carte de son de l'ordinateur permet de numériser la voix. La voix est découpée en portions réduites (quelques centièmes de secondes). L'ordinateur se charge ensuite de reconstruire ces portions en phonèmes, qui sont des éléments indivisibles du langage. Ensuite, ces phonèmes sont déterminés «acoustiquement» par traitement du signal sonore et par comparaison avec une bibliothèque de phonèmes stockés dans le système. Ce dernier va coller les phonèmes les uns aux autres pour retrouver les mots dictés[8]. »

Pour être performants, ces logiciels doivent tenir compte de trois grandes variables :

Une question de langue : La reconnaissance vocale nécessite l'identification de la langue parlée, car son fonctionnement repose sur l'analyse de phonèmes[9], qui sont différents d'une langue à l'autre. En français, il y a 37 phonèmes, en anglais, 42 et… plus de 400 en mandarin ! Lorsqu'un système de téléphonie nous demande si nous voulons utiliser le français ou l'anglais, ce n'est pas seulement pour savoir en quelle langue nous répondre, c'est aussi pour savoir en quelle langue nous « écouter ». Le système va également reconnaître les mots qu'il possède dans sa banque. Si un mot n'est pas connu, le système ira automatiquement au mot le plus près phonétiquement, à moins d'avoir été programmé pour réagir autrement.

Une question de son : Lorsqu'on ouvre un micro pour parler à un système de reconnaissance, le système « entend » tout ce que le micro peut capter, la voix aussi bien que le son ambiant. Un bruit ambiant trop fort est souvent à l'origine de problèmes de compréhension, car le système ne peut pas distinguer entre les voix et le bruit : tout n'est que du son. C'est la raison pour laquelle on suggère de parler à l'ordinateur avec un micro situé tout près de la bouche. Beaucoup de recherches sont en cours pour réduire l'impact du son ambiant, surtout pour les systèmes de reconnaissance dans les automobiles et pour les kiosques d'information interactifs dans les lieux publics.

Une question d'interlocuteur : Le changement d'interlocuteur peut également avoir un impact sur un système de reconnaissance vocale. Les systèmes publics, au vocabulaire restreint et bien défini, ont été entraînés pour de multiples interlocuteurs. Par contre, dans les équipements personnels (PC, téléphones, etc.), il faut entraîner l'équipement à reconnaître la façon de parler de son « maître », car le rythme, la prosodie, l'accent, la prononciation peuvent différer d'une personne à l'autre. L'entraînement de la machine à la reconnaissance de la voix n'est donc pas un système de sécurité, mais « l'apprentissage » nécessaire du logiciel.

---

[8] LE GROUPE ELMARZAK, DICAMILLO, CONTALDE. « La reconnaissance vocale », [en ligne]. www.geneve.ch/heg/campus/travaux/igs/sites/2002_04/Reconnaissance_Vocale.htm

[9] Phonème: «la plus petite unité de son (le pendant de morphème) pouvant être identifiée dans un flux de paroles et sémantiquement distincte». *L'ingénierie linguistique,* glossaire sur Internet

3.2 La reconnaissance de l'interlocuteur

Le système de reconnaissance de l'interlocuteur, plutôt que de reconnaître ce qui a été dit, vérifie si l'interlocuteur est bien la personne qu'il prétend être. La reconnaissance de l'interlocuteur fait partie de la biométrie, un ensemble de moyens techniques par lequel on peut identifier un individu. Nous sommes ici dans le domaine de la sécurité des individus, de la sécurité d'accès aux outils informatisés à la sécurité des transactions avec les institutions financières. Diverses applications impliquant la relation de l'homme avec la machine (ordinateur, automobile, système de sécurité, etc.) sont déjà en opération. En outre, des systèmes de reconnaissances de l'interlocuteur sont à l'essai dans le système pénal de certains pays pour s'assurer que le prisonnier libéré sur parole observe les termes de sa libération[10]. Par exemple, si une des conditions est que le prisonnier doit être chez lui à compter de 18 h le soir jusqu'au matin, le système l'appellera à des heures au hasard pour vérifier qu'il est bien là.

Une des méthodes pour reconnaître la voix d'un interlocuteur fonctionne de la façon suivante :

- Lors d'une tentative d'accès, on compare la voix d'un individu à la signature vocale gardée en mémoire. L'individu doit prononcer une phrase type et un logiciel sépare les différentes harmoniques de la voix et les compare à celles enregistrées.

- Les harmoniques sont au nombre de huit et définissent certaines fréquences privilégiées de la voix humaine. Ces sons purs ont des caractéristiques particulières (intensité, fréquence) dues à la forme des cordes vocales qui elles-mêmes sont suffisamment distinctes d'un individu à l'autre[11].

Les technologies de reconnaissance vocale font l'objet de nombreuses recherches pour des raisons de sécurité nationale, surtout depuis les événements du 11 septembre 2001. Elles suscitent par ailleurs certaines polémiques parmi les experts. Plusieurs considèrent que notre capacité de déchiffrer l'empreinte vocale n'a pas et ne devrait pas avoir de valeurs juridiques, à tout le moins dans les conditions actuelles. « *Or, contrairement aux idées reçues, la voix n'offre pas des caractéristiques individuelles aussi stables et fiables que celles des empreintes digitales utilisées par la police depuis un siècle, et aussi difficilement réfutables que celles des empreintes génétiques, une technique découverte en 1985. La voix n'est pas une image du corps comme le sont les photos des crêtes papillaires, et encore moins une représentation d'une partie du corps comme le sont les empreintes génétiques. En clair, dans l'état actuel des connaissances, il n'existe pas de procédures permettant d'avancer avec certitude qu'une personne est — ou n'est pas — l'auteur d'un appel téléphonique ou d'un enregistrement audio. Les rapports d'expertise d'enregistrements vocaux n'ont donc aucune validité scientifique.*

[10] MARKOWITZ, Judith. « Voice Biometrics – Are You Who You Say You Are? », *Speech Technology Magazine*, novembre/décembre 2003, [en ligne]. www.speechtechmag.com.

[11] « Biométrie et la vie privée », *Techno/parano,* [en ligne]. http://cyberzoide.developpez.com/opinions/index.php3?page=biometri.

La raison de notre inaptitude à attribuer un profil vocal à une voix donnée est la suivante : un enregistrement de parole n'est que la capture indirecte de mouvements articulatoires complexes faisant intervenir les cordes vocales, la langue, le voile du palais, la mâchoire et les lèvres. Les mouvements des organes de la parole engendrent des variations de pression acoustique instantanée qui peuvent être captées par un transducteur et transformées en variations de tension électrique. Or, comme tous les gestes de l'homme, les gestes de parole sont difficilement reproductibles à l'identique au cours du temps, sauf entraînement systématique. En effet, la vitesse d'articulation, l'intensité et la hauteur de notre voix varient beaucoup selon les conditions de communication (conversation familière, lecture, communication téléphonique, etc.), selon notre état psychologique et émotionnel, notre fatigue ou stress et, bien entendu, selon que nos cordes vocales et notre gorge se portent bien ou mal. La reconnaissance automatique de la parole, qui reste peu fiable encore aujourd'hui, est d'ailleurs directement confrontée à cette variabilité individuelle. »[12]

La reconnaissance vocale demeure donc à l'heure actuelle un chantier ouvert d'expérimentation, même s'il est possible d'y avoir recours concrètement pour certains usages bien définis et circonscrits.

4. La synthèse vocale

La synthèse vocale est la « reproduction, par une machine informatique, de la voix humaine à partir de données »[13]. Nous sommes ici dans le domaine des voix artificielles.

L'utilisation la plus connue de la synthèse vocale est la conversion de textes en paroles. En anglais, on utilise l'expression *Text-to-speech conversion* ou *TTS*.

Il existe plusieurs méthodes pour synthétiser la parole. À l'annexe B, on trouvera une liste non exhaustive de différentes méthodes de synthèse vocale présentement utilisées ou sous étude. La diversité de ces méthodes donne une bonne indication de la complexité du sujet et de l'immensité du champ de recherche.

Les fonctions d'un synthétiseur vocal dictent l'importance et la complexité de la base de données qu'on devra constituer. Le lecteur pour un handicapé visuel a besoin d'une base de données lourde. Il en va de même pour le système de réponse au service à la clientèle d'une compagnie.

Il semble que les meilleurs résultats de synthèse vocale proviennent d'un système dans lequel on a emmagasiné plusieurs heures d'enregistrements de « paroles » qui seront ensuite segmentées phonétiquement. Les échantillons doivent être nombreux, car lors des activités de « synthèse », il doit y avoir plusieurs exemples disponibles, chacun avec une prosodie différente[14].

Il va sans dire que la mise en place de ce synthétiseur est très laborieuse, très coûteuse et très lourde pour un système informatique. De plus, si le synthétiseur doit offrir plus d'une voix, le processus d'enregistrement et de segmentation doit être fait pour chaque voix.

---

[12] BOË, Louis-Jean. *Ben Laden et le mythe de l'empreinte vocale*, Institut de la Communication Parlée, INPG-Université Stendhal, [en ligne]. vivantinfo.com

[13] *Le Grand Dictionnaire de terminologie* de l'Office de la langue française

[14] DUTOIT Thierry, et autres. *Synthèse Vocale et Reconnaissance de la Parole: Droites Gauches et Mondes Parallèles*, Faculté Polytechnique de Mons, 2002, [en ligne]. tcts.fpms.ac.be/publications/papers/2002/cfa2002_tdlcfmvpcr.pdf -

Lorsqu'on vérifie les caractéristiques des synthétiseurs vocaux offerts sur le marché, on y trouve toujours une référence au nombre de voix disponibles.

Aujourd'hui, certains fournisseurs de synthétiseurs vocaux offrent la possibilité de se créer une image corporative, en se dotant d'une voix porte-parole, en se donnant un « son » qui désignera immédiatement l'entreprise — un peu comme le « son maison » des différents postes de radio ou de télévision.

Nortel, dans sa publicité sur Internet pour son système de *Text-to-Speech*, offre à ses clients la possibilité d'avoir la voix de leur porte-parole comme voix du système de téléphonie. On y indique qu'avec 20 à 40 heures d'enregistrement d'échantillons vocaux de la personne porte-parole et plusieurs semaines de segmentation, la « voix » sera prête à prononcer tous les messages et données requis[15].

Scansoft offre la même possibilité dans la publicité pour *Speechify TTS*, *Virtuoso V.I.P. (Voice Identity Program)*. ScanSoft offre un démo de ce système dans son site Internet[16]. Scansoft est la compagnie connue du public pour son logiciel *Dragon Naturally Speaking* destiné au PC.

Parmi les champs d'application de la synthèse vocale, on trouve actuellement :

- les systèmes de téléphonie;
- les services automatisés de service à la clientèle;
- les programmes d'apprentissage ou de perfectionnement personnel;
- les lecteurs automatisés pour les personnes souffrant d'un handicap visuel;
- les appareils téléphoniques pour les personnes souffrant d'un handicap auditif ou vocal.

À long terme, on laisse entendre dans la littérature disponible qu'on pourrait voir apparaître des « conteurs d'histoire à la demande » *(storyteller-on-demand)* ou encore des « visages parlants » *(talking-heads)*. Dans ces cas, les technologies de synthèse vocale et de reconnaissance vocale pourraient être associées à de l'animation visuelle pour produire de « l'audiovisuel interactif ». Ces produits audiovisuels interactifs pourraient servir à des fins éducatives aussi bien qu'à des fins de divertissement. Pour ces dernières applications, les systèmes ne sont toutefois pas encore au point.

La recherche dans le domaine de la synthèse vocale comme dans celui de la reconnaissance vocale continue. Les efforts visent d'abord à réduire la lourdeur des systèmes informatiques et non moindre défi, donner un son plus naturel aux voix, être en mesure de reproduire la prosodie du discours. En ce qui concerne ce dernier point, il y a peu de développements concluants au cours des dernières années. Rappelons qu'à l'heure actuelle, les logiciels arrivent à produire des voix plus naturelles que pour de courts segments du discours et, le plus souvent, pour des applications commerciales et industrielles.

[15] « With 20 to 40 hours of voice samples from the person whose voice is being captured and several weeks of tuning to create the voice model, the *voice* can be ready to speak all your automated prompts, messages and any other dynamic data required ». Tiré du site de Nortel Networks dans une description de produit intitulé *Why Text-to-Speech*, [en ligne]. www.nortelnetworks.com/products/04/eba/asr/doclib.html

[16] www.scansoft.com/speechify/customvoices

5. La conversion de la voix

5.1 Définition

La conversion de la voix est l'ensemble des techniques permettant de transformer les caractéristiques de la voix de façon à ce que les paroles dites par un interlocuteur X (source) semblent l'avoir été par l'interlocuteur Y (cible).

5.2 Les étapes de la conversion

La façon la plus reconnue à ce jour, dans l'élaboration de systèmes de conversion de la voix, se divise en trois étapes :

L'enregistrement des voix : il faut enregistrer la voix de l'interlocuteur X (source) et la voix de l'interlocuteur Y (cible) disant, chacune de leur côté, exactement le même échantillonnage de mots ou de phrases. Les échantillons doivent être dits dans la même langue et au même rythme.

L'analyse : le programme de conversion comparera les échantillons de Y avec ceux de X pour en tirer les éléments de comparaison. Ces éléments permettront d'établir le code de conversion (*codebook*) entre les deux voix. Il est important de retenir que tout système de conversion de la voix doit absolument avoir une étape d'analyse. C'est la base même des systèmes de conversion : découvrir la relation entre deux voix.

La conversion : en utilisant le code de conversion obtenu au cours de l'analyse, le système convertira un nouvel enregistrement de l'interlocuteur X de façon à faire croire que les nouvelles paroles ont été dites par l'interlocuteur Y.

Par exemple, des échantillons des voix de Peter et de Jean seront enregistrés et analysés par le système de conversion. Ensuite Peter enregistrera la phrase « Est-ce que tu connais ma mère ? ». Le système convertira la phrase « Est-ce que tu connais ma mère ? » de façon à nous faire croire que c'est Jean qui l'a dite.

Les aspects les plus perceptibles de la voix sont :

- le ton;
- l'intensité;
- la sonorité;
- le rythme.

Ces différents aspects peuvent dans une certaine mesure être traités de façon indépendante les uns des autres.[17] Donc il serait possible dans le processus de conversion d'utiliser le ton, l'intensité et la sonorité de l'interlocuteur Y tout en gardant le rythme de

[17] VERHELST, Werner, et Henk BROUCKZON DE VRIJE.. *Voice Modification for Lip Synchronization, Voice Dubbing & Karaoke*, Universiteit Brussel, novembre 2002, IEEE Benelux Workshop on Model Based Processing and Coding of Audio

l'interlocuteur X. En d'autres termes, en utilisant l'exemple mentionné plus haut, la phrase convertie, qu'on croit dite par Peter, aurait le ton, l'intensité et la sonorité de Peter, mais elle garderait le rythme de Jean, conservant ainsi le synchronisme avec la façon dont Jean a prononcé la phrase.

5.3 Les limites des systèmes de conversion existants

Le processus de conversion que nous venons de décrire est davantage une hypothèse de travail à l'heure actuelle car, en pratique, certains problèmes ou contraintes majeurs se sont manifestés :

Les échantillons identiques : il faut que les échantillons vocaux dits par l'interlocuteur X et l'interlocuteur Y soient identiques, non seulement en mots et dans la même langue, mais aussi dans la façon de les dire. S'il y a une différence dans le rythme, dans l'intonation, le code de conversion contiendra des éléments qui généreront de la distorsion[18].

La ressemblance phonétique entre la voix de l'interlocuteur X et l'interlocuteur Y : il faut que les voix de X et la voix de Y se ressemblent phonétiquement. Plus elles sont semblables, meilleur sera le résultat de la conversion. Plus elles sont différentes, plus la fréquence et la gravité de la distorsion augmenteront[19].

Le bilinguisme de X : dans le cas de doublage, il faut de plus que l'interprète doubleur, ici appelé *X*, parle couramment la langue de la version originale, donc qu'il soit bilingue. Il doit d'abord interpréter le personnage dans sa version originale, après quoi l'on sera en mesure d'établir le code de conversion avec la voix du comédien de cette version originale. L'interprète doubleur enregistrera par la suite la nouvelle version et on convertira cet enregistrement à l'aide du code de conversion établi. En procédant autrement, il y aura trop de différences entre le rythme, la prononciation et l'intonation des deux voix, ce qui générera des distorsions importantes. C'est donc un procédé qui s'avère très lourd.

L'imperfection de la ressemblance finale de la « nouvelle voix produite » avec celle de Y : Bien qu'il y ait une certaine ressemblance entre la voix convertie et la voix de Y, cette ressemblance n'est jamais parfaite, il demeure toujours une certaine distorsion.

Comme on peut le voir, les technologies de la conversion de la voix comme celles de la synthèse ne sont pas encore au point, à tout le moins pour des usages importants et de facture professionnelle. Si les recherches sont susceptibles d'apporter des solutions aux diverses problématiques soulevées, on peut dire d'ores et déjà que ces dernières divisent les systèmes de conversion des voix en deux grandes catégories : les systèmes de conversion des voix dans la même langue et les systèmes de conversion des voix d'une langue à une autre.

[18] *Voice Modification for Lip Synchronization, Voice Dubbing & Karaoke*
[19] *Voice Modification for Lip Synchronization, Voice Dubbing & Karaoke*

5.4 La conversion de voix dans la même langue

L'intérêt de la recherche dans ce domaine provient des applications concrètes qu'elles peuvent permettre. Voici quelques exemples d'applications possibles :

- la personnalisation des voix synthétiques qu'on retrouve dans les systèmes de téléphonie et dans les systèmes informatisés de service à la clientèle. Comme il a été mentionné ci-dessus, la création d'une voix dans le synthétiseur vocal est un long processus (et donc très coûteux) qui exige beaucoup de mémoire informatique. Le processus de conversion de la voix simplifierait ce processus et occuperait beaucoup moins d'espace dans l'ordinateur. La conversion pourrait ainsi constituer une alternative simple à la création de plusieurs voix pour toute forme de synthétiseur;
- l'amélioration des lecteurs automatisés pour les personnes souffrant d'un handicap visuel. Apparemment, les utilisateurs des lecteurs synthétiques peuvent se fatiguer de toujours entendre la même voix. La conversion de la voix pourrait offrir une alternative peu dispendieuse;
- le développement des lecteurs de courriel permettant l'écoute de courriels. Si l'empreinte de la voix est disponible, le courriel pourrait être lu avec la voix de la personne qui l'a envoyé;
- le « remplacement » de la voix de personnes qui prévoient la perdre à la suite de maladie. Si, avant de perdre la voix, une personne était en mesure d'enregistrer l'empreinte de sa voix, le synthétiseur qu'elle utiliserait, par la suite, pour « parler », pourrait le faire avec sa propre voix;
- la postsynchronisation dans la production originale d'œuvres audiovisuelles. Si un acteur de la version originale n'est pas disponible pour enregistrer la post-synchronisation, les enregistrements pourront être faits par un autre acteur et la voix de cet acteur sera convertie en la voix de l'acteur original;
- la possibilité de faire revivre des comédiens disparus. Si la technologie de l'animation permet de créer un personnage virtuel d'un comédien décédé, la re-création de « sa » voix viendra compléter la reconstitution. Il est à noter qu'en Californie, les droits sur la voix d'une personne prennent fin 70 ans après son décès[20];
- et bien sûr, comme une des multiples nouvelles armes au service du monde militaire. Il est bien évident qu'on ne dispose d'aucune information sur l'existence de recherches de ce côté. Une lecture de l'article paru dans le *Washington Post* du 1er février 1999 par William M. Arkin sous le titre *« When Seeing and Hearing Isn't Believing »*[21] suffirait à nous laisser imaginer à quelles fins la conversion de la voix pourrait être utilisée.

[20] *Califonia Civil Code*, article 3344 (g)
[21] Disponible dans Internet à l'adresse suivante :
www.washingtonpost.com/wp-srv/national/dotmil/arkin020199.htm

5.5 La conversion des voix d'une langue à une autre

Bien que plusieurs promoteurs prétendent disposer d'un système de conversion de la voix qui s'appliquerait d'une langue à une autre, les seuls démos disponibles dans Internet illustrent des exemples de conversion à l'intérieur d'une même langue.

On doit se rappeler que la conversion de voix d'une langue à une autre comporte des difficultés particulières, comme nous l'avons évoqué précédemment. En outre, lorsque la comparaison est faite entre deux voix, elle est faite entre des sons ou phonèmes comparables. Or, si le système doit les convertir en sons ou phonèmes pour lesquels il n'y a pas de correspondance établie, comment la conversion pourra-t-elle être faite ? En utilisant toujours le même exemple qu'au point 5.3, supposons que l'échantillonnage de Peter et Jean soit fait en anglais : comment, en français, se ferait la conversion de *tu* et de *mère* puisque les sons *u* de *tu* et *è* de *mère* n'existent pas en anglais ? Le système devrait-il les inventer ? Est-ce réaliste ? Et qu'en serait-il pour des langues comme le mandarin qui ne compte pas moins de 400 phonèmes ?

Nonobstant ces interrogations, certains chercheurs sont confiants que la conversion de la voix d'une langue à une autre pourrait devenir applicable pour les usages suivants:

- la traduction simultanée informatisée : un système de traduction simultanée informatisé et portatif, *Verbmobil*, a récemment été développé[22]. Un tel système donne la possibilité aux gens de langues différentes de parler et de se faire interpréter en temps réel. Par contre, si plusieurs personnes utilisent le même système lors d'une conférence, il est difficile de distinguer qui parle, car toutes les voix sont similaires. La possibilité de convertir la voix synthétique de l'interprète mobile en la voix de l'interlocuteur original serait fort utile[23]. La difficulté à surmonter ici, en plus de la question de passer à une autre langue, vient du fait qu'il n'y a pas d'analyse des empreintes de la voix de l'interlocuteur avant le moment de leur utilisation.

- le doublage : le comédien doubleur pourrait enregistrer le dialogue dans la langue du doublage, ensuite sa voix pourrait être convertie en la voix du comédien de la version originale selon le processus expliquée au point 5.3 en page 15.

Le lecteur remarquera ici que les standards de qualité à atteindre dans la conversion de voix peuvent varier d'une application à une autre. La conversion de voix en doublage exige une qualité supérieure à celle demandée dans un cadre de traduction simultanée.

---

[22] WAHLSTER, Wolfgang. *Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final Verbmobil system*

[23] SÜNDERMANN, David, et Hermann NEY. *An Automatic Segmentation and Mapping Approach for Voice Conversion Parameter Training,* Département d'informatique de RWTH Aachen – University of Technology, Allemagne

SÜNDERMANN, David, Hermann NEY et Harald HÖGE. *VTLN-Based Cross-Language voice Conversion*, Département d'informatique de RWTH Aachen – University of Technology, Allemagne et Siemens AG.

5.6 Les logiciels disponibles

Étant donné que la conversion de la voix est encore en période de recherche et de développement, il y a peu de logiciels disponibles. De plus, en raison des intérêts économiques en cause et du droit d'auteur, il s'avère très difficile d'obtenir de l'information sur des logiciels qui seraient soi-disant en développement ou sur le point d'être mis en marché.

Voici néanmoins quelques exemples de logiciels qui sont disponibles actuellement :

Le système *Voxonic* de la compagnie **SesTek,** en Turquie. SesTek est une jeune compagnie en recherche et développement qui se spécialise dans le développement des systèmes de téléphonie et des systèmes applicables à la parole. L'inventeur du programme est Oytun Turk. D'ailleurs *Voxonic*, sous le nom *Vox*, faisait partie de sa thèse de maîtrise en 2003[24]. On trouve deux exemples de conversion sur le site de Sestek[25], l'un du turc au turc et l'autre de l'anglais à l'anglais, sans indication du temps consacré à exécuter ces conversions.

Le logiciel *Vocal Imitation* de la compagnie **SisBit**, établie en Israël. Cette compagnie vient tout juste de mettre en marché ce logiciel qui permettrait divers traitement de voix dont la conversion de voix en doublage. Ce système se vend 130 USD. *Les essais de ce logiciel réalisés au Studio Place Royale (SPR Inc.) à Montréal ont démontré qu'il ne s'agissait pas d'équipement performant. La qualité des résultats est assurément insuffisante pour des activités de doublage de niveau professionnel (NDLR).*

6. Les impacts sur le doublage

6.1 La conversion des voix en doublage : des contraintes particulières

L'application des technologies de conversion des voix en doublage présente des exigences particulières dont les suivantes :

La production du code de conversion d'une voix vers une autre : selon les hypothèses actuelles, la conversion des voix en doublage nécessite que le comédien doubleur enregistre d'abord le texte dans la langue et tel que dit par le comédien de la version originale afin d'établir le code de conversion entre les deux voix. Cette étape constitue donc au départ une contrainte de temps et d'argent supplémentaire au processus de doublage actuel, sans compter qu'il se peut fort bien que le comédien doubleur ne parle pas cette langue. Il faudrait donc que le code de conversion puisse être composé à partir d'autre chose que des échantillons dits dans la même langue.

[24] *New Methods for Voice Conversion*, Bogaziçi University
[25] www.sestek.com.tr/english/sirket.html

La qualité du son : la qualité des systèmes de son dans les cinémas et celle des systèmes de cinéma maison exige des bandes sonores de qualité supérieure, mesurée en taux d'échantillonnage. Ainsi le taux d'échantillonnage exigé pour le doublage est de 44,1 kHz, mais seulement de 22 kHz pour les ordinateurs, de 11 kHz pour l'Internet et de 8 kHz pour les télécommunications. *La qualité des échantillons de conversion de voix que nous avons pu écouter n'atteint pas encore les standards établis pour le doublage. (N.D.L.R.)*

L'absence de toute forme de distorsion : il ne doit y avoir aucune distorsion dans les voix converties. La qualité de la voix obtenue doit être telle qu'elle puisse subir l'intervention des équipements périphériques sans effets défavorables.

L'automatisme : il faudrait que le processus de conversion des voix puisse être performant de façon entièrement automatisée. Si l'intervention humaine devient nécessaire pour corriger et peaufiner les résultats, des coûts supplémentaires s'ajouteront, sans compter que la réussite de la conversion deviendra alors une question subjective.

La ressemblance à la voix du comédien de la version originale : la voix convertie devra ressembler énormément à la voix du comédien de la version originale, sinon tout l'exercice s'avérera inutile et pourra même provoquer un agacement chez le spectateur.

6.2 Les impacts sur le processus de doublage d'une production

Le recours à un système de conversion des voix en doublage, tel que conçu actuellement, nécessitera des nouveaux outils ou étapes de travail dans le processus de doublage, notamment :

- l'enregistrement indépendant de la voix du comédien de la version originale. En effet, il faudrait un enregistrement dédié, car il est difficile d'extraire un échantillon « propre » de la voix du comédien à partir de la version originale du film. Rappelons ici que les voix ne sont pas isolées, mais qu'elles sont intégrées au mixage final du film;
- le transfert de l'échantillon de la voix du comédien de la version originale dans le système de conversion de la voix de la maison de doublage;
- l'enregistrement d'un échantillon (texte identique à celui du comédien de la version originale) par le comédien doubleur;
- l'analyse des échantillons précités afin d'en tirer le code de conversion;
- l'enregistrement du rôle complet du comédien doubleur séparément des autres comédiens, ce qui prolongera le temps de studio d'une part et augmentera le nombre de pistes à recaler et à mixer, d'autre part;
- la conversion de la voix du comédien doubleur en la voix de l'acteur original. Si cette conversion est faite au complet avant le mixage, cela implique du temps supplémentaire dans une salle dédiée à la conversion. Si cette conversion est faite pendant le mixage, cela implique qu'un logiciel de conversion soit disponible pour chacune des voix à être convertie.

On doit comprendre que l'ensemble des opérations que nous venons de décrire devra être appliqué pour chacun des rôles ou à tout le moins pour les voix qu'on aura décidé de

convertir. En outre, il ne sera plus possible d'enregistrer plusieurs rôles en même temps pour des « scènes collectives », chaque voix devant être enregistrée indépendamment. Cette pratique encourra des coûts supplémentaires de studio, de direction de plateau, de recalage et de mixage.

6.3 Les avantages et les contraintes du recours aux techniques de conversion des voix

Les raisons pour lesquelles on pourrait vouloir recourir à un système de conversion de la voix en doublage sont de trois ordres :

- conserver la voix du comédien de la version originale d'une production;
- être fidèle à une voix particulière;
- réduire les coûts et le nombre de comédiens requis pour un doublage.

Examinons les incidences de ces différents choix.

6.3.1 La conservation de la voix du comédien de la version originale

La volonté de conserver la voix d'un comédien de la version originale d'un film pourrait s'appuyer sur un souci de qualité et de fidélité plus grande à cette version ou parce qu'éventuellement un comédien d'une telle version pourrait exiger que sa voix soit utilisée lors des doublages de l'œuvre. Une telle décision a cependant des répercussions juridiques et financières non négligeables.

En premier lieu, ce choix nécessite que le comédien de la version originale donne la permission d'utiliser sa voix. La loi californienne[26] aussi bien que la loi québécoise[27] interdisent l'utilisation de la voix d'une personne sans sa permission.

De plus, outre le cachet que le comédien de la version originale pourrait demander pour l'utilisation de sa voix et du contrôle accru qu'il pourrait demander quant au choix du comédien doubleur, se pose la question de la sécurité. En effet, avec l'essor que prennent les technologies de reconnaissance de l'interlocuteur (voir point 3.2), le recours aux techniques de conversion de la voix pourrait être soumis à diverses contraintes pour « protéger » les empreintes vocales des individus.

Quant au comédien doubleur, sa voix ne serait pas reconnaissable dans le doublage, mais la performance de la voix doublée serait toujours la sienne. À moins d'avis contraire, la conversion pourrait être considérée comme un traitement similaire à ce qui est fait présentement en mixage avec les équipements périphériques (voir point 1.2.6).

[26] *California Civil Code,* Article 3344 (a)
[27] *Le Code civil du Québec*, Article 36 (5)

6.3.2 La fidélité à une voix particulière

La conversion de la voix pourrait être utilisée pour conserver les qualités particulières d'un personnage réel fort coloré, par exemple Darth Vader dans la série *Star Wars*, ou d'un personnage de dessin animé tel que Mickey Mouse, Pooh, Daffy Duck, etc. Dans ce cas, les technologies de conversion de la voix conféreraient au producteur ou au réalisateur d'une œuvre l'assurance que tel ou tel personnage conservera, peu importe la langue, la « même » voix.

Présentement, lors d'un doublage d'animation d'envergure, tous les efforts sont faits pour que la voix du doublage ressemble le plus possible à celle de la version originale. Il est même possible d'obtenir dans une version doublée une voix qui ressemble à s'y méprendre à la voix de la version originale. Ce résultat s'obtient en trouvant, par des auditions, une voix semblable à celle de la version originale. Une fois la nouvelle voix enregistrée, on fignole et peaufine la ressemblance à l'aide des équipements périphériques. Un système efficace de conversion de voix automatiserait une grande partie du travail. Quant à l'augmentation des coûts, on peut penser qu'elle serait marginale, puisque déjà, en doublage, les voix, qui requièrent un traitement spécial, sont toutes enregistrées sur des pistes séparées.

6.3.3 La réduction des coûts et du nombre de comédiens requis pour un doublage

Comme nous l'avons démontré précédemment, la conversion des voix augmente le nombre, le temps et les coûts des opérations de doublage, les coûts de libération de droits, ainsi que le temps studio, direction de plateau, recalage, mixage, etc. Les économies ne semblent pas au rendez-vous.

Quant aux cachets des comédiens, nous prenons pour acquis que les mêmes règles de calcul de cachet demeurent en place. Rappelons que les comédiens en doublage sont payés soit à la ligne, soit à l'heure, selon le plus élevé. Le nombre de lignes ne peut pas changer, donc le cachet le moindre serait le cachet calculé à la ligne.

Dans les cas de *doublage destiné à la télévision,* étant donné qu'il y a généralement peu de comédiens et qu'il est possible d'enregistrer plusieurs épisodes pendant la même session d'enregistrement, le plan d'enregistrement est organisé de façon à ce que le paiement des cachets soit le plus près possible du cachet à la ligne. Donc dans le cas de doublage pour la télévision, il n'y aura pas d'économie à utiliser moins de comédiens.

Dans le cas de *doublage pour le cinéma,* le temps accordé pour la réalisation d'un doublage par les distributeurs ou les producteurs est très court (il y a généralement deux codes de vitesse : urgent ou très urgent) et le travail se fait souvent sur une copie non finale du film. Les acteurs doublant des rôles principaux sont généralement payés à la ligne de texte. Donc, qu'un comédien double un rôle, deux rôles ou plus, cela ne change pas le cachet à payer pour le doublage de ces rôles. Dans le cas des petits rôles (et au cinéma, il y en a généralement beaucoup), les comédiens en doublage sont souvent payés à l'heure. Par contre, le coût d'échantillonnage et de conversion de la voix de tous ces

petits rôles viendra tout probablement contrebalancer ce qui pourrait être épargné en cumulant des rôles. De plus, il serait difficile dans les horaires serrés de doublage de trouver le temps de faire l'échantillonnage et la conversion. En cinéma, la conversion de la voix ne serait pas efficace.

Quant au cumul des rôles, l'*Entente collective* établit une limite au nombre de rôles qui peuvent être cumulés par un acteur. La limite n'est pas uniquement une question de cachet, c'est aussi une question de qualité. On ne choisit pas un comédien en doublage uniquement pour le son de sa voix. L'éventail de jeu, l'énergie, la passion entrent en ligne de compte dans le choix d'un comédien. Si un même comédien devait jouer beaucoup de rôles dans la même production, la qualité du doublage en souffrirait.

En conséquence de ce qui précède, il est clair que pour l'heure la conversion des voix n'engendre pas de réduction des coûts et du nombre de comédiens requis pour un doublage.

6.3.4 La conversion des voix en doublage : résumé des avantages et des impacts éventuels

Advenant que les technologies de conversion de la voix deviennent suffisamment performantes pour répondre au standard de qualité nécessaire aux œuvres de fiction (cinéma et télévision), elle pourrait permettre d'utiliser la voix des comédiens de la version originale d'une œuvre ou encore favoriser la création ou la reproduction d'une voix particulière (voir point 6.3.2).

Après analyse, nous croyons que le recours à la technologie de conversion de la voix pourrait avoir les impacts suivants au Québec :

- **Sur le processus et les coûts de production d'un doublage** :

  Si la conversion ne vise qu'une voix particulière, les coûts de production d'un doublage ne devraient pas augmenter de façon sensible. Toutefois, si la conversion vise l'ensemble des voix d'une version originale ou plusieurs rôles, il y aurait augmentation des coûts pour couvrir les enregistrements séparés des comédiens, l'enregistrement des échantillons, le temps de studio supplémentaire requis pour les conversions, le recalage et le mixage, etc.

- **Sur les entreprises de doublage** :

  L'entreprise de doublage devrait se procurer l'équipement et les logiciels requis pour la conversion des voix et s'assurer que le(s) comédien(s) de la version originale ont donné leur consentement à ce que leur voix soit utilisée pour le doublage de la production.

  Si l'obtention de l'autorisation d'utiliser la voix d'un comédien de la version originale devient complexe pour le producteur/distributeur de la production, ceci pourrait éventuellement devenir une raison supplémentaire de ne produire qu'un seul doublage en français — soit la version provenant de la France. Il pourrait

donc y avoir un impact éventuel sur le chiffre d'affaires des entreprises de doublage et sur le volume de travail des comédiens, mais pour l'heure, cela demeure de l'ordre de la supposition.

- **Sur le travail des comédiens** :

  Les comédiens retenus pour l'interprétation d'un rôle devront parler couramment la langue de la version originale de la production devant être doublée et devront enregistrer le texte dans sa version originale (ou autre échantillon significatif) pour que soit d'abord établi le code de conversion des voix. Autrement, leur travail demeurera sensiblement le même. Par ailleurs, s'il advenait que la conversion des voix favorise dans certains cas la production d'un seul doublage en français, il pourrait peut-être y avoir un impact sur le volume de travail comme nous l'avons supposé ci-dessus.

- **Sur les producteurs et les distributeurs des œuvres :**

  Ils devront s'assurer que les comédiens de la version originale ont donné leur autorisation pour l'utilisation de leur voix et que tout autre clause y afférant soit respectée. Enfin, ils devront assumer les coûts supplémentaires d'un doublage avec conversion des voix. Le producteur ou le distributeur qui commande le travail de doublage devra également fournir à la maison de doublage les échantillons de voix nécessaires pour la conversion.

7. En conclusion

Les technologies de conversion des voix ne sont pas encore au point. Par ailleurs, comme elles pourraient être utiles à toutes sortes d'applications, il y a fort à parier qu'elles progresseront dans l'avenir et que les recherches en cours aboutiront à des outils adaptés aux différentes utilisations qu'on pourra en faire, y compris en doublage.

Pour l'heure, on doit par ailleurs constater que la conversion des voix en doublage demeure une hypothèse de travail et advenant que des outils efficaces soient rendus disponibles, notre analyse nous porte à croire qu'elle ne deviendra pas nécessairement une pratique automatique en doublage. Des nombreuses dimensions juridiques, financières et de temps sont en cause. Il se pourrait bien que les différents intervenants n'aient recours à cette technologie que pour des cas particuliers.

À ce stade-ci de développement de la technologie de la conversion de la voix en doublage, nous n'avons pas cru pertinent d'enquêter auprès des *majors* pour connaître leurs intentions. Advenant des développements concrets, on pourra réévaluer la pertinence de communiquer avec eux pour connaître leurs intentions à cet effet.

8. Les modes de suivi du dossier

Les recherches dans le domaine de l'ingénierie linguistique sont nombreuses. Pour faire un monitoring efficace, nous croyons qu'il faut suivre les travaux des associations internationales reconnues par les milieux scientifiques dans ce domaine, soit l'ICSLP[28], IEEE[29] et Eurospeech. L'ICSLP est un des forums internationaux bien cotés pour la présentation de nouveaux développements dans le monde de l'ingénierie linguistique. Les conférences se tiennent aux deux ans. La prochaine réunion aura lieu en octobre 2004 en Corée. À ce jour, on ne semble pas prévoir y traiter de la conversion de la voix. Mais, au cours d'une réunion du *IEEE International Conference on Acoustics, Speed and Signal Processing (ICASSP 2004)* prévue à Montréal en mai 2004, il y aura une présentation sur la conversion de la voix.[30] Tant qu'à Eurospeech, il s'agit d'une association internationale à but non lucratif qui regroupe des spécialistes des technologies de la langue parlée. Cette association tient également des colloques importants à tous les deux ans.

Une révision périodique des dossiers présentés par ces associations serait une bonne indication de l'état de la recherche. Ces organismes rendent public sur Internet le contenu de leurs rencontres, y compris les documents qui y circulent. On trouve également sur Internet certaines revues qui sont spécialisées dans les développements en ingénierie linguistique. *Speech Technology Magazine* en est un exemple. La version Internet de la revue est disponible gratuitement et peut donc être consultée aisément. Advenant des développements réels et pertinents à notre dossier, il y a tout lieu de croire que cette revue s'en ferait l'écho.

****

[28] International Conference on Spoken Language Procession

[29] The IEEE (Eye-triple-E) is a non-profit, technical professional association of more than 360,000 individual members in approximately 175 countries. The full name is the Institute of Electrical and Electronics Engineers, inc., although the organization is most popularly known and referred to by the letters I E E E www.ieee.org/portal/index.jsp

[30] *Voice Conversion and Morphing Algorithms for TTS Systems* [en ligne]. www.icassp2004.com/Papers/viewpapers.asp?papernum=1276

Annexe A — Les changements affectant la bande rythmo

La bande rythmo est un ruban 35 mm clair sur lequel est écrit à l'encre de chine le texte du doublage. C'est cette bande qui sert à la projection en studio.

Le ruban 35 mm utilisé est de l'amorce claire en acétate. L'amorce claire en acétate devient de plus en plus rare, aussi la remplace-t-on progressivement par de l'amorce claire en polyester. Or, l'encre de chine n'adhère pas bien à cette surface, ce qui nuit à la bonne marche du travail de doublage. Les entreprises de doublage sont donc à la recherche de solutions de rechange depuis un bon moment.

Récemment, deux nouveaux procédés ont été mis au point pour solutionner ce problème. Le premier « Rythmique » est un système informatique français tandis que le deuxième « Dub Studio » a été conçu par une entreprise québécoise, Ryshco média. Bien que les deux systèmes semblent proposer des alternatives à la bande rythmo traditionnelle, ils sont configurés de façon sensiblement différente.

« Rythmique » est un système informatique qui fonctionne en synchronisme avec une image projetée sous format DVD. Il propose une bande rythmo numérisée. Le système transpose les opérations faites traditionnellement sur l'amorce en opérations faites sur une surface informatique, et sur laquelle on écrit au crayon optique et à la main. Le détecteur fait son travail sur le niveau de travail détection, l'adaptateur change le niveau et écrit l'adaptation, le calligraphe change encore le niveau et écrit la calligraphie. Enfin, le dactylographe fait son travail dans une fenêtre de texte qui apparaît sur le même écran que la bande rythmo. Le tout est très similaire à ce qui se fait maintenant. Les artisans se sentent très à l'aise. Il est à noter que le texte écrit à la main dans ce système n'est pas reconnu comme du texte par l'informatique et ne peut donc être traité comme tel.

« Dub Studio » est différent. L'image est numérisée. Le système reconnaît les changements de plans et les identifie sur la bande avec indication visuelle du code temporel. Le texte doit être intégré au logiciel, reproduit au complet et en conformité avec certaines spécifications. En utilisant la reconnaissance vocale, le système écoute le dialogue et place le texte à l'endroit approprié sur la bande. Un détecteur vérifie le travail du système et l'ajuste au besoin en utilisant la souris ou en ajoutant du texte à l'aide du clavier. L'adaptateur écrit son texte de doublage au clavier et le texte se place par-dessus ou au-dessous de la détection. L'adaptateur peut ajuster la synchro en utilisant la souris. Quant à la dactylographie, le texte est déjà dans le système; il suffit en principe de l'imprimer, mais il serait sans doute important de le valider avant de l'imprimer et surtout, avant d'aller en studio. Le système génère aussi une grille et un décompte de lignes. On remarquera ici qu'avec « Dub Studio », les étapes de calligraphie et de dactylographie, telles que décrites dans le processus traditionnel du doublage (point 1.2.3 du rapport) sont supprimées.

En ce qui a trait à l'ingénierie linguistique, « Dub Studio » utilise la reconnaissance vocale pour « placer » le texte original. Il utilise des connaissances de traitement de texte pour faire certaines opérations et permet, par exemple, de remplacer des mots par des abréviations lorsqu'il n'y a pas de place sur la bande (*i.e.* : « vous » par « vs » etc.).

De façon plus générale, ce système présente une architecture ouverte et de nouveaux outils pourraient y être intégrés de manière à améliorer ses diverses fonctions. Il demande toutefois une plus grande adaptation des pratiques de la part des artisans du doublage.

On notera que « Dub Studio » est présentement utilisé dans un studio de doublage à Montréal pour des émissions de télévision. Il n'a toutefois pas encore subi l'épreuve des doublages « urgents » et « très urgents » des longs métrages pour le cinéma.

Les deux systèmes possèdent des points forts et des points faibles. Il reste à voir quel système l'industrie favorisera !

Annexe B — Quelques méthodes de synthèse vocale

La synthèse articulatoire

Technique par laquelle on tente de produire des paroles en imitant le système articulatoire humain. Il est possible que cette forme de synthèse donne des résultats parmi les plus satisfaisants, mais pour le moment la recherche et l'expérimentation ne sont pas suffisamment avancées. Cette forme de synthèse requiert une compréhension très précise du fonctionnement de tout le système articulatoire.

Les définitions présentées ci-dessous sont tirées du Grand dictionnaire technologique de l'Office québécois de la langue française.

La synthèse par formants

Technique de synthèse vocale qui simule les fréquences de résonance du conduit vocal humain à l'aide de filtres. Les formants sont des composantes de la parole qui permettent de distinguer un son complexe d'un autre. Ils correspondent aux fréquences (ou les bandes de fréquences) les plus intenses; on les obtient en faisant l'analyse du son.

La synthèse de la parole à partir du texte

Technique de synthèse vocale qui consiste à transformer le texte en message vocal à la suite d'une conversion des éléments du texte, les graphèmes, en éléments sonores, les phonèmes.

La synthèse par diphonèmes

Technique de synthèse vocale qui consiste à reproduire des sons à partir de diphonèmes, c'est-à-dire des éléments sonores composés du dernier segment d'un son et du premier segment du son suivant (l'équivalent anglais *synthesis by diphones* est tiré de *Computer Speech Processing,* 1985, p. 64).

Les éléments mémorisés dans ce système correspondent aux caractéristiques d'une langue donnée. Les diphones sont extraits de l'enregistrement d'un locuteur ou d'une locutrice donné. La constitution du dictionnaire nécessite une phase d'analyse (prédiction linéaire ou formants) et une phase de segmentation semi-automatique (extraction des diphones) qui peut prendre entre une semaine et un mois. Le dictionnaire correspond aux caractéristiques vocales d'un locuteur donné. Plusieurs dictionnaires ont déjà été élaborés pour la synthèse du français (André Abbou, Thierry Meyer et Isabelle Lefaucheur, *Les industries de la langue. Les applications industrielles du traitement de la langue par les machines*, vol. 2, 1987, p. 76). La synthèse par diphonèmes est réalisée à l'aide d'un dictionnaire contenant environ 1 200 diphonèmes.

La synthèse par des mots

Technique de synthèse vocale qui consiste à faire prononcer des messages vocaux par un système automatique à partir de mots préalablement enregistrés et gardés en mémoire. La synthèse par mots est très coûteuse et ne permet de synthétiser qu'un vocabulaire limité.

La synthèse par prédiction linéaire

Technique de synthèse vocale qui fonctionne sur le principe de la prédiction linéaire, c'est-à-dire qui peut prévoir, dans une très grande mesure, la valeur d'un échantillon sonore, prélevé à un instant donné, à partir de la valeur des dix ou douze échantillons qui l'ont précédé. Les équivalents anglais *linear predictive synthesis* et *LPC synthesis* sont donnés dans *Electronic Speech Synthesis. Techniques, Technology and Applications,* 1984, p. 70.

La synthèse par règles

Technique de synthèse vocale qui procède à partir d'un ensemble de règles définies, comme les règles de transition d'un phonème à l'autre, de prononciation de rythme et d'accent, qui régissent chacun des paramètres de la production d'un message vocal. Les équivalents anglais *synthesis-by-rule* et *rule synthesis* sont donnés dans R. Linggard, *Electronic Synthesis of Speech,* 1985, p. 15. La synthèse par règles a été mise au point par D. Klatt du *Massachusetts Institute of Technology* (d'après Jean Guilbert et Michel Guilbert, *Les ordinateurs qui parlent,* 1986, p. 125).

La synthèse par restitution de parole compressée

Technique de synthèse vocale qui consiste à enregistrer un ensemble de caractéristiques des éléments de parole (phonèmes, syllabes ou mots) d'un vocabulaire donné, à les stocker en mémoire et à les reproduire au besoin.

Annexe C — Liste des documents consultés

L'ingénierie linguistique — général

« Ingénierie du Langage — Exploiter toutes les ressources du langage » [en ligne].
[www.hltcentral.org/usr_docs/Whatis/Whatis-fr.htm]

« Ingénierie linguistique » [en ligne]. [www.culture.gouv.fr/culture/dglf/riofil/enjeux.htm]
« Ingénierie linguistique – la reconnaissances vocale » [en ligne].
[www.culture.gouv.fr/culture/dglf/riofil/recon-vocal.htm]

« Ingénierie linguistique – les enjeux » [en ligne].
[www.culture.gouv.fr/culture/dglf/riofil/lang-naturel.htm]

« Ingénierie linguistique – les outils » [en ligne]
[www.culture.gouv.fr/culture/dglf/riofil/outils.htm]

« Ingénierie linguistique – les ressources : les corpus » [en ligne].
[www.culture.gouv.fr/culture/dglf/riofil/ressources-linguistiques.htm]

« L'Ingénierie linguistique ou comment exploiter la puissance du langage » [en ligne].
[www.hltcentral.org/usr_docs/Harness/harness-fr.htm]

« Qu'est-ce que l'ingénierie linguistique » [en ligne].
[http://zeus.fltr.ucl.ac.be/autres_entites/LING/GELI2DC/definition.htm]

Les unités linguistiques : phonèmes, segments, etc.

« English Phonemes, Spelling and Meaningful Representations » [en ligne].
[www.auburn.edu/~murraba/spellings.html]

« The Idea of segments » [en ligne].
[www.umanitoba.ca/faculties/arts/linguistics/russell/138/sec3/segment.htm]

« Liste des phonèmes du français » [en ligne].
[http://retore.chez.tiscali.fr/LPC/phoneme.htm]

« Phonèmes et graphies du français » [en ligne].
[http://talana.linguist.jussieu.fr/~weini/LG_00-01/IPA_fr.html]

« Spectrogram Reading; Spectral Cues for the Broad Categories of Speech sounds » [en ligne].
[http://cslu.cse.ogi.edu/tutordemos/SpectrogramReading/ipa/ipadefault.html]

« Spectrogram Reading; Spectral Cues for Phonemes » [en ligne].
[http://cslu.cse.ogi.edu/tutordemos/SpectrogramReading/ipa/ipachars.html]

« Spectrogram Reading; What are Formants? » [en ligne].
[http://cslu.cse.ogi.edu/tutordemos/SpectrogramReading/ipa/formants.html]

La Reconnaissance Vocale et la Synthèse Vocale

DUTOIT, Thierry, et autres. «Synthèse Vocale et Reconnaisance de la Parole : Droites Gauches et Mondes Parallèles » FACULTÉ POLYTECHNIQUE DE MONS [en ligne].
[http://tcts.fpms.ac.be/publications/papers/2002/cfa2002_tdlcfmvpcr.pdf]

GROUPE ELMARZAK, DICAMILLO, CONTALDO. « La reconnaissance vocale » [en ligne].
[www.geneve.ch/heg/campus/travaux/igs/sites/2002_04/Reconnaissance_Vocale.htm]

« Ingénierie linguistique – la reconnaissances vocale » [en ligne].
[www.culture.gouv.fr/culture/dglf/riofil/recon-vocal.htm]

LEMMETTY, Sami. « A Review of Speech Synthesis Technology », UNIVERSITY OF TECHNOLOGY, Helsinki [en ligne].
[www.acoustics.hut.fi/~slemmett/dippa/thesis.pdf]

MARKOWITZ, Judith. [www.speechtechmag.com/]
« Voice Ideas : Truth in Advertising », *Speech Technology Magazine,* [en ligne]. (march/april 2002)
« Hocus Pocus » *Speech Technology Magazine,* [en ligne]. (july/august 2002)
« The For-Real Story » *Speech Technology Magazine,* [en ligne]. (november/december 2002)
« Voice Biometrics – Are You Who You Say You Are? » *Speech Technology Magazine,* [en ligne]. (november/december 2003)

« Synthèse Vocale PAGS/SINOLA » [en ligne].
[www.ircam.fr/produits/technologies/pags.html]

La conversion de la voix

ARKIN, William M. « When Seeing and Hearing Isn't Believing » *Washington Post*, 1er février 1999, [en ligne].
[www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A45085-2000Feb28]

CEYSSENS, Tim, Werner VERHILST et Patrtick WAMBACQ. « On the Construction of a Pitch Conversion System », CENTRE FOR PROCESSING SPEECH AND IMAGES DEPT. FOR ELECTRICAL ENGINEERING KATHOLIEKE UNIVERSITEIT » [en ligne].
[www.etro.vub.ac.be/Research/DSSP/publications/int_conf/EUSIPCO-2002.pdf]

CHAO WANG, Ming Tang et Stephanie SENEFF. « Voice transformations : From Speech Synthesis to Mammalian Vocalizations » EUROSPEECH 2001, Aalborg, Danemark, [en ligne]. [www.sls.csail.mit.edu/sls/publications/2001/phase_vocoder.pdf]

EZZAT, Tony, Jim GLASS et T. POGGIO. « Audio Morphing », [en ligne]. [www.ai.mit.edu/projects/cbcl/res-area/abstracts/2004-abstracts/ezzat-am.pdf]

GUTIÉRREZ-ARRIOLA, J.M., et autres. « A New Multi-speaker Formant Synthesizer that applies Voice Conversion Techniques », DPTO. DE INGENIERIA ELECTRONICA, ETSIT, UNIVERSIDAD POLITECNICA DE MADRID, Eurospeech 2001, [en ligne]. [www-gth.die.upm.es/research/documentation/articuloJuanaEurospeech2001.PDF]

HOUCHTARIS, Athanasios, Jan VAN DER SPIEGEL et Paul MUELLER. « Non-Parallel Training for Voice Conversion by Maximum Likelihood Constrained Adaptation » UNIVERSITY OF PENNSYLVANIA et CORTICON INC., SYMPOSIUM DE L'ICASSP 2004, [en ligne]. [www.icassp2004.com/Papers/viewpapers.asp?papernum=1276]

HUANG, Ying, et autres. « Real-time Lip Synchronization based on Hidden Markov Models » DEPT. OF ELECTRICAL ENGINEERING TSINGHUA UNIVERSITY, MICROSOFT RESEARCH, CHINA

LIN, Cheng-Yuan, et J.-S. ROGER JANG. « New Refinement Schemes for Voice Conversion » DEPT. OF COMPUTER SCIENCE, NATIONAL TSING HUA UNIVERSITY, TAIWAN, IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA & EXPO, BALTIMORE – juillet 2003, [en ligne]. [www.icme2003.com/Papers/viewpapers.asp?papernum=1820]

ORPHANIDOU, Christina, Irene M. MOROZ et Stephen J. ROBERTS. « Voice Morphing Using the Generative Topographie Mapping », OXFORD CENTRE FOR INDUSTRIAL AND APPLIED MATHEMATICS ROBOTICS RESEARCH GROUP, DEPT OF ENGINERRING SCIENCE, UNIVERSITY OF OXFORD, [en ligne]. [www.maths.ox.ac.uk/~orphanid/orph_CCC03.pdf]

PUTERBAUCH, John. « Voice Conversion » PRINCETON UNIVERSITY, [en ligne]. (http://silvertone.princeton.edu/~john/voiceconversion.htm)

SAKAMOTO, Masaharu et Takashi SAITO. « Speaker Recognizability Evaluation of a Voicefont-Based Text-to-Speech System » IBM RESEARCH, TOKYO RESEARCH LABORATORY ICSLP 2002, [en ligne]. [http://webtts.watson.ibm.com/pubs/icslp2002-saka.pdf]

SCHULTZ T., et A. WAIBEL. « Multilingual and Crosslingual Speech Recognition » INTERACTIVE SYSTEMS LABORATORIES, [en ligne]. [www.informedia.cs.cmu.edu/mli/papers/darpa-Broadcast-ws98-tanja.pdf]

TURK, Oytun. « New Methods for Voice Conversion » par Oytun » BOGAZICI UNIVERSITY [en ligne]. [www.busim.ee.boun.edu.tr/~speech/thesis/oytun_turk.pdf]

TURK, Oytun, et Levent M. ARSLAN. « Subband Based Voice Conversion » [en ligne]. [www.busim.ee.boun.edu.tr/~speech/publications/Voice_Conversion/Subband_Based_Voice_Conversion.pdf]

VERHELST, Werner, et Henk BROUCKZON. « Rejection Phenomena in Inter-Signal Voice Transplantations » VRIJE UNIVERISTEIT, BRUSSELLES, IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, New Paltz, NY – octobre 2003, [en ligne]. [www.etro.vub.ac.be/Research/DSSP/publications/int_conf/WASPAA-2003.pdf]

VERHELST, Werner et Henk BROUCKXON. « Voice Modification of Lip Synchronization, Voice Dubbing and Karaoke » IEEE BENELUX WORKSHOP ON MODEL BASED PROCESSING AND CODING OF AUDIO (MPCA-2002) [en ligne]. [www.etro.vub.ac.be/Research/DSSP/publications/loc_conf/MCPA-2002-A.pdf]

VERHELST, Werner, Tim CEYSSENS et Patrick WAMBACK. « On Inter-Signal Transplantation of Voice Characteristics » IEEE BENELUX SIGNAL PROCESSING SYMPOSIUM (SPS-2002), [en ligne]. [www.etro.vub.ac.be/Research/DSSP/publications/loc_conf/SPS-2002-A.pdf]

YE, Hui. « High Quality Voice Morphing » CAMBRIDGE UNIVERSITY ENGINEERING DEPARTMENT, [en ligne]. [http://mi.eng.cam.ac.uk/~hy216/VoiceMorphingSeminar.pdf] (à noter des exemples de conversion de la voix)

Le système VERBMOBIL

RINSCHEID, Ansgar. « Voice Conversion Based on Topological Feature Maps and Time-Variant Filtering » LEHRSTUHL FUR ALLGEMEINE ELEKTROTECHNIK UND AKUSTIK, RUHR-UNIVERSITAT BOCHUM, ALLEMAGNE [en ligne]. [www.asel.udel.edu/icslp/cdrom/vol3/235/a235.pdf]

SUNDERMANN, David, et Hermann NEY. « An Automatic Segmentation and Mapping Approach for Voice Conversion Parameter Training », COMPUTER SCIENCE DEPT, RWTH AACHEN –UNIVERSITY OF TECHNOLOGY, ALLEMAGNE, [en ligne]. [www-i6.informatik.rwth-aachen.de/PostScript/InterneArbeiten/Suendermann_AST_2003.pdf]

SUNDERMANN, David, et Hermann NEY. « VTLV-Based Cross-LanguageVoice Conversion », COMPUTER SCIENCE DEPT, RWTH AACHEN –UNIVERSITY OF TECHNOLOGY, ALLEMAGNE. [en ligne]. [www.i6.informatik.rwthaachen.de/PostScript/InterneArbeiten/Suendermann_ASRU_2003.pdf]

SUNDERMANN, David, et Hermann NEY. « VTLV-Based Voice Conversion » COMPUTER SCIENCE DEPT, RWTH AACHEN –UNIVERSITY OF TECHNOLOGY, ALLEMAGNE, [en ligne].[wwwi6.informatik.rwthaachen.de/PostScript/InterneArbeiten/Suendermann_ISSPIT_2003.pdf]

WAHLSTER, Wolfgnang. « Mobile Speech-to-Speech Translation of Spontaneous Dialogs : An Overview of the Final Verbmobil System », [en ligne]. [http://verbmobil.dfki.de/ww.html]

L'ingénierie linguistique — divers

ARPPE, Antti. « Forward with Feet on the Ground – Speech Technology the Finnish Way » FINNISH IT CENTER FOR SCIENCE [en ligne]. [www.csc.fi/euromap/artikkelit/puheteknologia.phtml.en]

Le grand dictionnaire terminologique de l'Office québécois de la langue française [en ligne]. http://w3.granddictionnaire.com/btml/fra/r_motclef/index800_1.asp

Liste des sujets à l'ordre du jour de la rencontre de l'ICASSP du 18 mai 2004 concernant la conversion de la voix [en ligne]. [www.icassp2004.com/Papers/PublicSessionIndex3.asp?Sessionid=1046]

THOM, D., H. PURNHAGEN et MPEG AUDIO SUBGROUP DE L'ORGANISATION INTERNATIONALE DE NORMALISATION. « MPEG Audio FAQ Version MPEG-4 », [en ligne]. www.chiariglione.org/mpeg/faq/mp4-aud/mp4-aud.htm

Sequentia – vol II, n° 4, juin/juillet/août 1995

La dimension juridique

California Civil Code, article 3344 concerning right of publicity, [en ligne]. [www.leginfo.ca.gov/cgi-bin/waisgate?WAISdocID=11925014523+0+0+0&WAISaction=retrieve]

Code Civil du Québec, Articles 35 et 36

Independent Production Agreement entre l'ACTRA, le CFTPA et L'APFTQ

La cause de Bret Michaels, c. Internet Entertainement group, inc et al., [en ligne]. [www.kentlaw.edu/classes/rwarner/legalaspects/michaels_jurisdiction.html]

L'Entente collective entre l'Union des artistes et l'Association des Doubleurs Professionnels du Québec (1er mars 2003 au 28 février 2006)

Les références à des entreprises et à des produits

Norbec Communication pour le système *Rytmique*

Nortel Networks concernant la synthèse de voix, [en ligne].
[www.nortelnetworks.com/products/04/eba/asr/doclib.html]
[http://a560.g.akamai.net/7/560/5107/20030925230353/www.nortelnetworks.com/products/04/oscar/collateral/nn103943-081203.pdf]

Ryshco Media pour le système *dub studio*

Scansoft pour les produits *Speechify*, [en ligne].
[www.scansoft.com/speechify/customvoices]

SesTek pour leur programme *Vox* [en ligne].
[www.sestek.com.tr/voice_conversion/demonstrations_eng.htm]

SisBit pour leur programme *Voice Imitation*. [en ligne]
[www.sisbit.com/Category.asp?cx=136&cid=94&x=136]

Annexe D — Liste des personnes rencontrées ou interviewées

BÉLANGER, Normand. — détecteur

BERGUA, Daniel. — TECHNICOLOR

BOULIANNE, Gilles. — (B. Ing. UNIVERSITÉ DE QUÉBEC À CHICOUTIMI et M.Sc INRS-Télécommunications) conseiller au CENTRE DE RECHERCHE INFORMATIQUE DE MONTRÉAL

BROSSEAU, Julie. — (M.A. en linguistique, UNIVERSITÉ DE MONTRÉAL), conseillère au CENTRE DE RECHERCHE INFORMATIQUE DE MONTRÉAL

CHÉNIER, Guylaine. — TECHNICOLOR

CÔTÉ, Jocelyne. — RYSHCO MÉDIA

FERNANDEZ, Gavin. — mixeur

JACQUES, Maud. — détecteure

JONES, Michael. — anciennement de WALT DISNEY CHARACTER VOICES, Los Angeles

O'SAUGHNESSY, Douglas. — professeur à L'UNIVERSITÉ DE QUÉBEC À MONTRÉAL

RODRIQUES, Normand. — et le personnel du STUDIO PLACE ROYALE

RYSHPAN, Howard. — RYSHCO MÉDIA

Le personnel du studio LES VILAINS GARÇONS

L'équipe de NORBEC COMMUNICATION pour la présentation du système de détection *Rytmique*