UN SYNTHÉTISEUR DE LA VOIX CHANTÉE BASÉ SUR MBROLA POUR LE MANDARIN

Liu Ning
CICM (Centre de recherche en Informatique et Création Musicale)
Université Paris VIII, MSH Paris Nord
liuningchine@yahoo.fr

RESUME

Dans cet article, nous présentons le projet de développement d'un synthétiseur de la voix chantée basé sur MBROLA en mandarin pour la langue chinoise. Notre objectif vise à développer un synthétiseur qui puisse fonctionner en temps réel, qui soit capable de se synchroniser avec un séquenceur MIDI ou puisse être piloté par un clavier MIDI pour produire la voix chantée en mandarin. Le développement de l'application est basé sur un algorithme existant - MBROLA, ainsi l'objet « mbrola~ » est utilisé dans notre programmation. Notre travail consiste en la création de la première base de données en mandarin pour MBROLA, et le développement d'une application informatique musicale dans l'environnement Max 5.

1. INTRODUCTION

Aujourd'hui, les technologies de la synthèse de la voix parlée sont très diversifiées et avancées, et en même temps les synthétiseurs pour la voix chantée sont en cours de développement. Les synthèses sont basées sur des méthodes différentes, par exemple la synthèse FM de John Chowning [1]. Le MUSSE (Music and Singing Synthesis Equipment) de Johan Sundberg est un synthétiseur qui a intégré la synthèse par règle et la synthèse vocale formantique [2]. Le programme CHANT de Xavier Rodet est basé sur la synthèse formantique [3]. Le SPASM (Singing Physical Articulatory Synthesis Model) est développé par Perry Raymond Cook basé sur le modèle physique [4]. Le logiciel Vocaloid est développé par l'université Pompeu Fabra, et commercialisé par YAMAHA, il est basé sur la synthèse sinusoïdale [5].

En 2002, le synthétiseur de la voix chantée « On-The-Fly » a été proposé par le département de l'informatique de Tsing Hua Université à TAIWAN [6]. L'algorithme de PSOLA a été employé pour mettre en œuvre cette synthèse.

En 2004, le professeur GU Hong Yan et son collègue CHEN An Rui de l'Université Nationale de Science et de Technologie de Taiwan ont développé et amélioré une synthèse additive basée sur la méthode sinusoïdale. Ce système est capable de produire la voix chantée en fonction de la mélodie et des paroles à partir d'un fichier MIDI [7].

Néanmoins, les synthèses de la voix chantée pour la langue chinoise ont encore des limites techniques. Par exemple, le système à partir de la lecture de fichier MIDI est un système en temps différé. La concaténation de l'unité enregistrée basée sur le phonème donne une dis-naturalité à la voix. Les synthèses ne produisent pas le changement de la fréquence fondamentale linéaire (utile pour le *glissando* ou le *portamento*).

Conscients de ces limites constatées dans les synthétiseurs de la voix chantée pour la langue chinoise existants, et dans la perspective d'apporter des améliorations, nous allons parler dans cet article de notre développement d'une application basée sur le synthétiseur MBROLA via « mbrola~ ». Cette application sera réalisée pour une synthèse en temps réel. Les fonctions de la production de la voix *portamento* et de la voix *glissando* seront intégrées dans cette application. Dans un premier temps, nous présenterons la création d'une base de données diphonique en mandarin pour MBROLA ; ensuite nous présenterons la programmation de l'application dans l'environnement Max 5.

2. L'ALGORITHME MBROLA ET L'OBJET « mbrola~ » POUR MAX 5

MBROLA (Multi-Band Re-synthesis Overlap Add) [8] est un algorithme connu de synthèse vocale basé sur la synthèse concaténative de l'unité diphonique. Le synthétiseur MBROLA a l'avantage de pouvoir produire une voix très intelligible. Grâce aux travaux des équipes de chercheurs de différents pays, aujourd'hui, le système MBROLA est capable de produire la parole pour 35 langues différentes.

En 2005, Nicolas D'Alessandro, Raphäel Sebbe, Baris Bozkurt et Thierry Dutoit du Laboratoire TCTS de la Faculté Polytechnique de Mons (Belgique) ont développé l'objet « MaxMBROLA~ » basé sur le synthétiseur MBROLA pour l'environnement programmation Max/MSP [9]. L'objet « MaxMBROLA~ » et «mbrola~ » (la nouvelle version de l'objet « mbrola~ »

est sortie en 2010) permet de rendre le fonctionnement du synthétiseur MBROLA en temps réel. « mbrola~ » a été développé pour Max 5. Nous pouvons utiliser cet objet pour charger une base de données MBROLA, modifier la durée originale de la séquence de la parole, varier la fréquence fondamentale de la voix, etc.

3. L'ENVIRONNEMENT MAX 5

Conçu par l'IRCAM et Cycling'74, Max 5 est destiné aux artistes, aux musiciens, aux designers sonores, aux enseignants ou encore aux chercheurs. Ce langage de programmation visuelle, proche de la programmation orientée objet donne l'accès à des nombreux modules et fonctions prédéfinies, et à des instructions pouvant être stockées pour faciliter l'assemblage au sein d'une interface graphique. Max 5 permet à l'utilisateur de faire des manipulations interactives, grâce au calcul et au traitement du signal audio en temps réel [10].

4. CRÉATION DE LA PREMIÈRE BASE DE DONNÉES MBROLA POUR LE MANDARIN

Pour la création de la base de données MBROLA, Nous devons d'abord trouver toutes les combinaisons diphoniques possibles et nécessaires pour la voix chantée en mandarin. Notre édition de la liste des diphones est principalement basée sur les règles de transcription phonétique PIN YIN [11]. 410 segments diphoniques étaient sélectionnés pour la base de données de la voix parlée. Nous savons que dans le chant, nous pouvons répéter la dernière voyelle pour prolonger une syllabe chinoise. Cependant, certaines voyelles ou combinaisons de voyelles ne sont pas présentées dans le dictionnaire chinois. Par exemple, dans notre figure1 ci-dessous, la dernière syllabe est prolongée par une prononciation /ei/, alors que cette prononciation propre à la voix chantée n'existe pas dans le dictionnaire pour la voix parlée. C'est la raison pour laquelle, après avoir effectué des analyses phonétiques et modifié certaines règles de prononciation de la voix parlée pour les adapter à la voix chantée, nous avons complété la base de données de la voix parlée en y ajoutant un certain nombre de prononciations qui n'existent que dans la voix chantée. Grâce à ces démarches, la liste des diphones de la voix parlée comprend finalement 519 segments au lieu de 410.

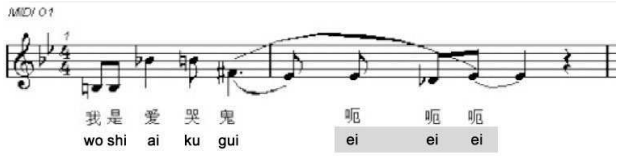

Figure 1. Extrait de chanson « Ai Ku Gui »

Le prélèvement des échantillons de la voix a été fait avec des matériels audio professionnels, la qualité originale de prélèvement de la voix est en 16bit 44.1kHz.

Pour avoir des prononciations parfaites de toutes les syllabes chinoises, deux chanteuses dont le caractéristique de la voix est proche ont participé dans les enregistrements. Lors de l'enregistrement, nous avons veillé à ce que la hauteur des prononciations soit constante. Après la MBROLISATION, le timbre de la voix est très acceptable.

Dans la phase de segmentation, nous avons bouclé les segments de voyelles afin de garantir la continuité dans le chant. Après beaucoup d'essais, nous avons également déterminé les durées générales des segments : la durée de l'initiale est environ 60ms et la durée de la finale est 200ms. Si dans le synthétiseur ces règles de durée sont respectées, la production de la syllabe chinoise peut être clairement audible, et le temps de réaction du changement de fréquence fondamentale du synthétiseur peut être réduit au delà du 200ms. La première base de données - « cn1 » de MBROLA pour la voix chantée et parlée de la langue chinoise a vu le jour.

5. DÉVELOPPEMENT DE L'APPLICATION

Nous avons conçu deux moyens différents pour contrôler le synthétiseur de voix chantée : la synchronisation du synthétiseur par un séquenceur MIDI pour l'utilisation en post-production ; ou bien une fonction qui nous permet de connecter directement le synthétiseur à un clavier MIDI pour une interprétation musicale en directe.

Quatre modules principaux sont inclus dans l'application (figure 2). Le module de synchronisation est développé pour recevoir les messages MIDI, il peut utiliser les messages MIDI pour piloter le séquenceur de la parole et la production de la voix. Le module de séquenceur de la parole est destiné à toutes les éditions de la parole, nous pouvons ajouter, modifier ou supprimer la parole dans une séquence, et la séquence de parole peut être enregistrée dans un fichier de texte après l'édition. Le module de règles gère les informations musicales, il peut convertir le format du message MIDI au format du synthétiseur MBROLA, par exemple la conversion de la hauteur de note ou encore les messages « note on/off ». Dans le module de la synthèse MBROLA, nous avons d'abord mis en place une bibliothèque de transcription phonétique, elle peut traduire la transcription PIN YIN en segments pour la base de données « cn1 ». Ensuite, nous avons développé un synthétiseur en temps réel basé sur l'objet « mbrola~ » pour la voix chantée chinoise.

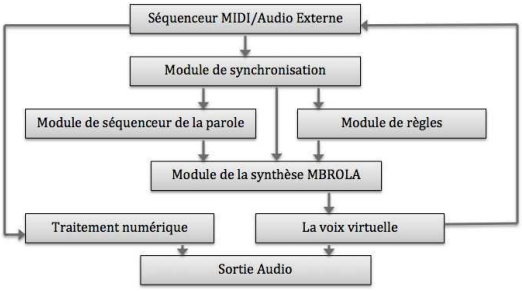

Figure 2. La structure de l'application

6. APPLICATION « MANDARIN REALTIME SINGING SYNTHESIS »

Cette application peut fonctionner en esclave en s'associant avec une autre application ou matériel séquenceur MIDI. Elle peut fonctionner de façon autonome pour une utilisation de l'interprétation *Live*.

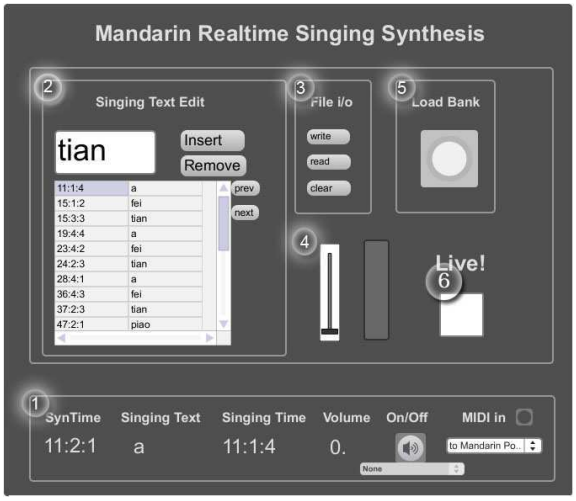

Figure 3. L'interface de l'application

L'interface d'application dispose de plusieurs modules (figure 3). En bas de la fenêtre, c'est le module de transmission ; nous voyons ici les informations du code de pointeur de Position, la parole en exécution, le code SMTPE, le volume, le menu pour la sortie audio, le bouton pour activer ou désactiver le synthétiseur MBROLA, le menu pour choisir le port MIDI, et un diode pour visualiser le signal d'entrée MIDI (figure 3, n°1).

Dans la partie en haut à gauche de l'interface, nous avons le module d'édition de la parole. Nous y trouverons le séquenceur de la parole, nous pouvons saisir, modifier ou supprimer la parole dans la séquence, les boutons « prev » « next » nous permettent d'avancer ou de reculer la position sur la séquence (figure 3, n°2).

Le module de « File i/o » est en haut dans le centre de l'interface, il inclut trois boutons – « write » « read » « clear ». Nous pouvons cliquer sur le bouton « write » pour sauvegarder la séquence de la parole dans un ficher texte externe. Nous pouvons charger une séquence de parole existante en cliquant le bouton « read ». Le bouton « clear » peut supprimer toutes les paroles dans la séquence actuelle (figure 3, n°3).

Au dessous du module « File i/o », nous avons un potentiomètre pour régler le volume général de la sortie audio, et à côté, nous avons un mètre de l'intensité du volume (figure 3, n°4).

En haut à droite de l'interface, nous avons le module « Load Bank », nous pouvons cliquer sur le bouton pour charger une base de données de MBROLA. Par exemple la base de données « cn1 » (figure 3, n°5).

Au dessous du module n°5, nous avons une case pour activer ou désactiver le mode *Live*. Une fois qu'il est activé, le synthétiseur va ignorer la synchronisation avec le séquenceur MIDI. La séquence de la parole s'enchaînera automatiquement après la réception de chaque message « note off » (figure 3, n°6).

Cette application est capable de produire la variation de hauteur de la voix linéaire. Par exemple, si nous produisons une syllabe /ma/ par une série de segments (figure 4), nous pouvons faire un *glissando* ou la voix *portamento* en modifiant la fréquence fondamentale de chaque segment comme la courbe rouge dans la figure 5. Au contraire, un changement non linéaire de la hauteur peut économiser le nombre de segment dans la production (veuillez-voir la courbe bleue dans la figure 5).

Figure 4. Une série de phonèmes enchaînés pour la prononciation /ma/

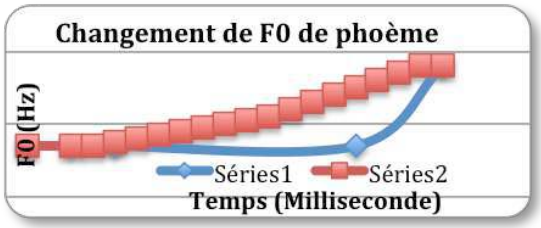

Figure 5. Changement de la fréquence fondamentale des phonèmes

Dans l'application « Mandarin Realtime Singing Synthesis », nous avons également intégré les fonctions de vibrato, et de *Pitch Bend*. Le synthétiseur peut répondre aux message de contrôle MIDI standard. Ces fonctionnalités peuvent améliorer la qualité de la voix chantée. La compatibilité de protocole MIDI nous

permet d'utiliser cette application avec de nombreux logiciels et matériels.

7. CONCLUSION

Dans cet article, nous avons fait une brève description du processus de développement de « Mandarin Realtime Singing Synthesis ». C'est le premier synthétiseur de la voix chantée basé sur une base de données diphonique en temps réel pour la langue chinoise. Cette application peut être utilisée dans le studio ou dans le concert. Les résultats obtenus de notre recherche pourront intéresser les musiciens et les chercheurs. De plus, La base de données « cn1 » que nous avons créée pendant le développement peut s'utiliser dans de nombreuses applications compatibles à MBROLA.

Néanmoins, notre recherche pourra être complétée et améliorée par des travaux ultérieurs, comme par exemples, l'amélioration de la latence du synthétiseur et la qualité de la production de la voix ; les possibilités de contrôle pour le traitement de la voix ; la diversification et spécification de la base de données en mandarin en créant des versions de voix d'enfant, ou masculine ; la réécriture de l'application par une langage de programmation de bas niveau ; la compatibilité de notre application avec les 34 autres langues de MBROLA, etc.

REFERENCES

[1] Cheng-Yuan Lin, J.-S. Roger Jang, Shaw-Hwa Hwang, "An On-the-Fly Mandarin Singing Voice Synthesis system", Proceedings of the Third IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing, Taiwan, 2003.

[2] Gu Hong Yan, Chen An Rui, "la méthode de la synthèse de la voix chantée en mandarin basée sur le modèle sinusoïde", Proceedings of the Conférence sur l'intelligence artificielle et les applications, Taiwan, 2004.

[3] Chowning John, "Frequency Modulation Synthesis of the Singing Voice", Proceedings of the Some Current Directions in Computer Music Research, MIT Press, Cambridge, UK, 1989.

[4] Johan Sundberg, "The KTH Synthesis of Singing", Proceedings of the Advances in Cognitive Psychology, Vol.2 No.2-3 p.133, Warsaw, Poland, 2006.

[5] Xavier Rodet, "Time domain formant-wave-function synthesis". pp.9-14, Computer music journal, 8(3), 1984.

[6] Cook Perry Raymond, "Spasm : a real-time vocal tract physical model editor/controller and singer : the companion software system", Proceedings of the Colloque sur les modèles physiques dans l'analyse, la production et la création sonore, Grenoble, France, 1990.

[7] Hideki Kenmochi, Hayato ohshita, "Vocaloid – Commercial singing synthesizer based on sample concatenation", Proceedings of the Centrer for Advenced Sound Technologies, Yamaha Corporation, Japan, 2007.

[8] T. Dutoit and H. Leich, "MBR-PSOLA : Text-to-Speech Synthesis Based on an MBE Resynthesis of the Segments Database", Proceedings of the Speech Communication, no°13, pp.435-440, Elsevier Publisher, USA, 1993.

[9] Nicolas D'Alessandro, Sebbe Raphaël, Bozkurt Baris, Dutoit Thierry, "Max/MSP Mbrola-Based Tool for reel-time voice synthesis", Proceedings of the EUSPIPC 05 Conference, Antalya ,Turkey, 2005.

[10] Louis Frécon, Okba Kazar, *Manuel d'intelligence artificielle*, p.554, PPUR Presses polytechniques, Lausanne, Suisse, 2009.

[11] Paul R. Frommer, Edward Finegan, *Looking at Languages: A Workbook in Elementary Linguistics*, p.365, Harcourt Brace College Publishers, San Diego, California, 1999.