# Un logiciel pour l'enseignement de la prosodie

Philippe Martin
LLF, UFRL, Université Paris Diderot Sorbonne Paris Cité
Place Paul Ricœur, 75013 Paris, France
philippe.martin@linguist.univ-paris-diderot.fr

## RESUME

On présente un logiciel de visualisation en temps réel de la mélodie de la phrase, WinPitch LTL, dont la dernière version est optimisée pour une utilisation pédagogique. Reprenant la disposition traditionnelle modèle-imitation sur deux fenêtres séparées, ce logiciel est équipé de fonctions qui permettent de répondre aux insuffisances souvent constatées dans d'autres réalisations. Celles-ci portent essentiellement sur l'absence d'indications visuelles claires sur les courbes mélodiques qui rendraient compte des éléments pertinents au regard d'une approche théorique convaincante et explicite. Une fonction d'alignement automatique permet alors de reporter sur la courbe de l'apprenant les segments mélodiques jugés pertinents, conduisant ainsi celui-ci à juger de sa performance. D'autres fonctions du logiciel comme le play-back au ralenti et le morphing prosodique en renforcent les aspects pédagogiques.

## ABSTRACT

**A visualizer to teach sentence prosody**
This paper presents a visualization software of sentence intonation operating in real time, WinPitch LTL, whose latest version is optimized for educational use. Taking the traditional model-imitation layout on two separate windows, this software has features that may meet the shortcomings often found in comparable realizations. These shortcomings mainly pertain to the absence of clear visual indications on the displayed melodic curves that would be related to some explicit theoretical approach. An automatic aligner maps model selected segments, deemed relevant, to the learner melodic segments, leading the learner to judge her/his performance. Other functions like slow playback and prosodic morphing reinforce the educational aspects of the software.

MOTS-CLES : intonation, prosodie, enseignement de l'oral, structure prosodique incrémentale.
KEYWORDS: intonation, prosody, language teaching, incremental prosodic structure.

# 1. Introduction

La visualisation de l'intonation à des fins pédagogiques, réalisé en temps réel ou différé, remonte au moins à 1964 (Vardanian, 1964). Beaucoup d'autres réalisations ont été proposées depuis (entre autres Léon et Martin, 1972 ; Abberton et Fourcin, 1975 ; Bot, 1980) avec des résultats pratiques discutables en ce qui concerne l'enseignement oral d'une langue seconde. Une des raisons expliquant leur possible inefficacité semble découler de l'absence de tout contexte théorique qui guiderait les apprenants, auxquels il était souvent demandé d'imiter globalement une courbe mélodique sans autre explication (James, 1976). De nombreux utilisateurs ont été rapidement découragés lorsque

l'acceptabilité de leurs réalisations n'a pas été évaluée par rapport à un objectif intonatif clair et un modèle phonologique ou phonétique explicite.

Pour être plus efficace, il apparait que la visualisation devrait être accompagnée d'une part d'explications précises quant aux réalisations prosodiques souhaitées, et de l'autre de la localisation automatique sur la courbe mélodique de l'apprenant des segments mélodiques admis comme pertinents. Cette pertinence, qu'elle soit phonologique ou phonétique, doit découler d'un modèle rendant compte des hauteurs ou des mouvements mélodiques spécifiques survenant sur telle ou telle syllabe, et non de caractéristiques globales de la courbe mélodique (durée, hauteur moyenne, etc.) (Martin, 1975, 2009). L'existence implicite d'un modèle est particulièrement importante compte tenu des variations acceptables de locuteurs natifs. Sinon, le système pourrait refuser une réalisation de l'apprenant, parfaitement acceptable pour un natif, mais dont le modèle et l'évaluation n'aurait pas tenu compte.

## 2. Des améliorations possibles

Les réalisations antérieures de visualiseur de mélodie à but d'enseignement de l'oral semblent donc, pour la plupart, souffrir de deux défauts : 1) le manque de modèle implicite pout guider l'apprenant dans ses réalisations prosodiques et 2) l'absence d'un système de localisation des segments considérés comme fautifs relativement au modèle dans la réalisation de l'apprenant. Ce dernier point doit permettre d'attirer l'attention sur des corrections ciblées et non globales. De plus, pour surmonter l'appréhension de certaines catégories d'apprenants, un tel système devrait pouvoir afficher, en plus de courbes mélodiques modèle et imitée, courbes relativement complexes qui peuvent désorienter les utilisateurs, des messages précis relativement aux erreurs constatées par le système. La visualisation deviendrait alors optionnelle.

WinPitch LTL, le logiciel présenté ici, a été spécialement conçu pour répondre à ces deux exigences en intégrant clairement sur les courbes mélodiques affichées des caractéristiques phonologiques jugées importantes par un expert auteur, lequel est libre d'élaborer des ensembles de phrases regroupées en leçons selon le modèle théorique choisi.

Les deux améliorations potentielles sont donc : a) un modèle explicite rendant compte des variations mélodiques (et donc de la structure prosodique des énoncés) dont les caractéristiques pertinentes sont affichées sur les courbes mélodiques et b) une fonction permettant de retrouver automatiquement les segments pertinents dans les réalisations de l'apprenant malgré les différences possibles de débit et de rythme par rapport au modèle.

## 3. Un modèle prosodique

L'ensemble du système étant ouvert grâce à de nombreuses fonctions auteur permettant d'élaborer des leçons d'exercices annotés au gré de l'instructeur, tout modèle prosodique peut convenir, pourvu qu'il soit aisément compréhensible par l'apprenant et que l'instructeur puisse en traduire les éléments pertinents sur les courbes mélodiques des énoncés modèles choisis. À côté des modèles phonologiques dominants, en particulier ceux dérivés de l'approche autosegmentale-métrique (Jun & Fougeron, 2002), on rappelle ici les grandes lignes d'un modèle théorique centré sur le caractère dynamique sur l'axe du temps des évènements prosodiques, tant du point de vue du locuteur que de celui de l'auditeur. Ce caractère dynamique implique entre autres que les évènements prosodiques, ont un caractère local et non global, et que leurs réalisations sous forme de contours mélodiques dépendent en général de leur contexte immédiat.

Du point de vue théorique choisi, il s'agit pour l'élaboration des exemples de marquer par différentes couleurs selon leur nature les contours mélodiques placés sur les syllabes accentuées et

singulièrement sur les voyelles accentuées. Ces syllabes sont considérées par hypothèse comme seules pertinentes pour l'indication de la structure prosodique de l'énoncé.

Le modèle de structure prosodique incrémentale est basé sur un principe clairement défini, l'indication d'une relation de dépendance "à droite" entre les événements prosodiques instanciés par des contours mélodiques placés sur les syllabes accentuées des groupes accentuels. Ces groupes se trouvent en français en position finale de syntagmes intonatifs. À partir d'un inventaire des classes de contours mélodiques en termes de relations de dépendance, une description simple de la structure prosodique émerge à partir du principe de contraste de pente.

Ce principe spécifie qu'un contour mélodique Cx indique une relation de dépendance envers un autre contour Cy "à droite" (i.e. apparaissant plus tard dans l'énoncé) par une pente mélodique de sens opposée. Ainsi un contour C1, souvent appelé de continuation majeure, indique une relation de dépendance envers un contour conclusif terminal déclaratif C0 par une variation mélodique montante, opposée à celle, descendante, du contour terminal. Ces relations de dépendance déterminent les regroupements successifs des unités prosodiques minimales que sont les groupes accentuels, c'est-à-dire les séquences de syllabes terminées par une syllabe accentuée (hors accent emphatique ou d'insistance).

On aboutit ainsi à une liste de contours mélodiques utilisés en français pour l'indication de la structure prosodique d'un énoncé.

Si C0, C1, C2 et Cn désignent ces contours mélodiques réalisés sur les syllabes (les voyelles) effectivement accentuées de l'énoncé, on a :

C0 contour descendant et bas ; C1 contour montant ; C2 contour descendant ; Cn contour neutralisé, de faible variation mélodique.

Pour attirer l'attention de l'apprenant sur leur pertinence, on peut surligner en couleurs différentes chacune de ces classes de contours. Par exemple, C0 rouge, C1 bleu, C2 marron, Cn noir, etc. au gré de l'auteur des exemples.

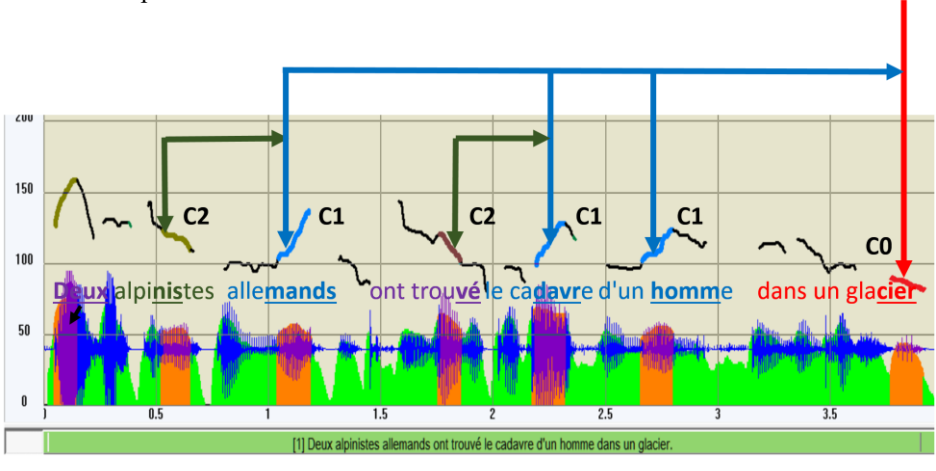


FIGURE 1 : Exemple d'un réseau de relations de dépendance indiquées par les flèches horizontales, chacun des contours pertinents du point de vue de l'approche théorique choisie par l'auteur étant colorée dans différentes couleurs (ici C0 rouge, C1 bleu, C2 marron). Les regroupements successifs des groupes accentuels sont indiqués par le parenthésage *[[deux alpi**nist**es C2][alle**mand**s C1]][[ont trou**vé** C2][le ca**davr**e C1]][d'un **homm**e C1][dans un gla**cier** C0]*

Les relations de dépendance indiquées par les contours sont résumées par l'expression Cn -> C2 -> C1 -> C0. Ces relations sont transitives, c'est-à-dire que Cn -> C2 implique aussi Cn -> C1 et Cn -> C0 (les flèches Cx -> Cy indiquent la dépendance de Cx par rapport à Cy). Une séquence C2 C0 est extrêmement rare et procède d'une marque d'insistance (Martin, 2009).

Les relations de dépendance opèrent « à droite » relativement au futur de la séquence de contours mélodiques. Ainsi un contour C1 de continuation majeure, en indiquant non seulement que l'énoncé n'est pas terminé, présuppose l'apparition d'un contour terminal C0 signalant la fin de l'énoncé et permettant à l'auditeur de rassembler les différents groupements de syntagmes pour accéder au sens de l'énoncé. L'existence de C1 dépend donc de l'apparition future d'un contour C0. Il en va de même pour le contour descendant C2, qui dépend de l'apparition future d'un contour C1, etc.

On notera que ce mécanisme d'indication de la structure prosodique par regroupements successifs de groupes accentuels va clairement à l'encontre du cadre théorique dominant basé sur la phonologie métrique-autosegmentale. En niant par hypothèse le rôle des accents mélodiques (pitch accents), cette approche ne peut évidemment rendre compte d'une quelconque interaction entre les contours mélodiques. Le cadre autosegmental-métrique ne propose en fait aucune explication quant à la fonction de la structure prosodique autre qu'actualiser la syntaxe, en se basant sur le concept du « bien-formé » emprunté à l'approche générative transformationnelle. Or il est relativement facile de démontrer que la structure prosodique n'est pas dérivée de la syntaxe (si ce n'est dans une certaine mesure dans l'oralisation de l'écrit), mais qu'au contraire elle la précède aussi bien pour le locuteur que pour l'auditeur (Martin, 2015).

# 4. Alignement

L'alignement des réalisations imitées par rapport au modèle est réalisé selon une procédure classique de déformation temporelle dynamique (algorithme DTW, Dynamic Time Warping). Son efficacité résulte du choix des fonctions de comparaison qui détermine un appariement optimal entre deux séries temporelles constituées par les énoncés à aligner. Ces séries temporelles sont déformées par transformation non-linéaire de la variable temporelle pour déterminer une mesure de leur similarité. Ce processus s'applique donc parfaitement au problème de la comparaison entre deux prononciations d'un même énoncé, modèle et imitation.

L'implémentation s'opère selon les étapes suivantes :

a. Un spectrogramme FFT à large bande est stocké dans une matrice temps-fréquence, dont les dimensions typiques sont 512 (nombre de créneaux de fréquences) par 2000 (nombre de valeurs de temps correspondant à un nombre standard de pixels horizontaux de l'écran).

b. Parallèlement, un spectrogramme LPC est stocké dans la matrice temps-fréquence. L'avantage de LPC par rapport à la FFT est lié à la possibilité d'utiliser une fenêtre de signal plus petite, et donc obtenir une meilleure résolution temporelle dans la comparaison des spectres relatifs au modèle et à l'imitation.

c. La matrice spectrographique (FFT ou LPC) est divisée en canaux de fréquence, de 500 Hz largeur de bande. La gamme de fréquence retenue est de 200 Hz (pour éviter le bruit de fond de l'enregistrement) à 4500 Hz (fréquence spectrale maximale utile pour la parole, à l'exclusion de la [s]). La répartition des bandes de fréquences est donc linéaire et non pas logarithmique (cette dernière, pourtant standard, s'étant avérée moins performante dans cette implémentation.)

d. Intensité : des courbes d'intensité sont calculées et stockées pour chaque canal.

Les comparaisons de spectre pour chaque déplacement sur la matrice fréquence-temps se font par rapport a ces valeurs d'intensité déjà calculées, permettant un calcul rapide.

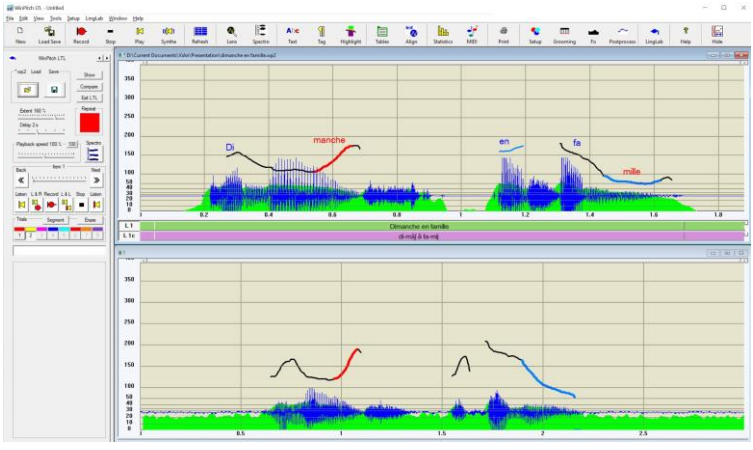


FIGURE 2 : Un exemple d'alignement automatique de segments censés être pertinents et surlignés sur la courbe mélodique modèle sur la courbe de l'apprenant. En rouge le contour de continuation C1, en bleu le contour conclusif terminal C0.

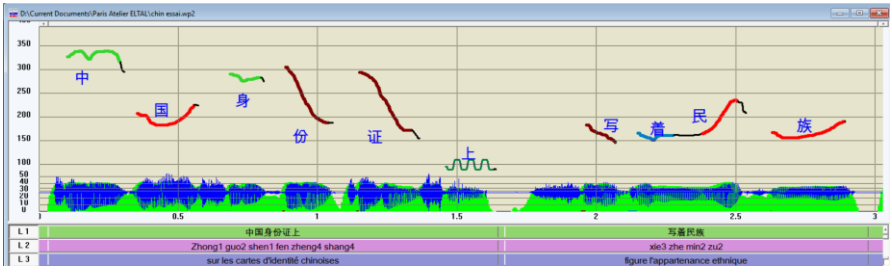


FIGURE 3 : Un exemple de surlignage couleur selon le ton d'une phrase en mandarin (Ton 1 vert, Ton 2 rouge, Ton 3 bleu, Ton 4 marron).

## 5. Ralenti

WinPitch LTL dispose également d'une fonction de ralenti. Basée sur la méthode TD-PSOLA (Moulines & Charpentier, 1990) et la détection des marqueurs de pitch par la méthode du peigne spectral (Martin, 1981), elle opère dans une gamme de 15% à 100% ce qui permet à l'apprenant de corréler visuellement la déplacement d'un curseur le long de la courbe mélodique avec le son de parole correspondant. En pratique, cette fonction s'avère très importante pour sensibiliser les utilisateurs non musiciens aux mouvements mélodiques, dont ils sont en général peu conscients. Le simple fait de pouvoir corréler le déplacement d'un curseur le long d'une courbe mélodique avec le segment de parole reproduit au ralenti sans déformation spectrale s'avère tout à fait convaincant pour mieux appréhender les variations mélodiques.

# 6. Morphing prosodique

Une autre fonction du logiciel WinPitch LTL permet de réaliser un morphing prosodique avec segmentation automatique intégrée, de manière à permettre le morphing des durées (rythmique) en plus du morphing mélodique. Cette fonction opère par segmentation automatique alignée sur les transitions spectrales (Martin et Delais, 2016) ou constante (une balise toutes les 100 ms par exemple) de manière à réaliser le morphing des durées pseudo-segmentales. Les variations mélodiques sont ensuite appliquées sur les segments correspondant du modèle à l'imitation de l'apprenant. Cette fonction permet de sensibiliser l'apprenant sur une réalisation proche du modèle utilisant sa propre voix, rendant psychologiquement ce but plus accessible.

# 7. Interface utilisateur

L'interface utilisateur dans sa version pour apprenant est représenté Fig. 4. Les différentes commandes et fonctions sont :

a. Lecture d'un fichier leçon, comprenant les énoncés proposés à l'apprenant.
b. Sauvegarde du fichier leçon, accompagné des différentes réalisations de l'apprenant (avec un maximum de 8 imitations par énoncé proposé).
c. Un réglage de l'échelle temporelle pour la visualisation de la courbe mélodique en temps réel.
d. Un réglage de la vitesse de playback du modèle et de l'imitation de l'apprenant.
e. Affichage d'une notice explicative en format HTML relative au modèle proposé.
f. Alignement et comparaison des segments mélodiques de l'apprenant par rapport au modèle.
g. Sélection du délai donné à l'apprenant pour démarrer son imitation (avec affichage du délai par couleurs rouge, orange et verte…).
h. Navigation dans le fichier leçon (sélection de l'énoncé précédent, suivant, ou par numéro d'ordre).
i. Sélection et affichage des réalisations précédentes de l'apprenant (maximum de 8).
j. Morphing prosodique automatique du modèle sur la réalisation de l'apprenant.

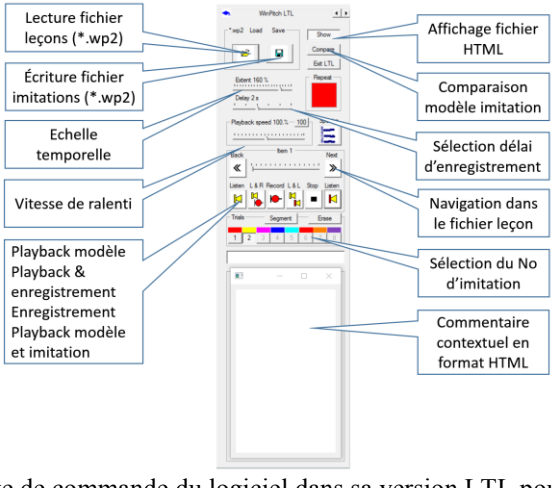


FIGURE 4 : Boîte de commande du logiciel dans sa version LTL pour apprenants

Le fichier de sauvegarde comprend les fichiers modèles accompagnés des différentes réalisations de l'apprenant, ce qui permet une évaluation et l'insertion de commentaires par l'instructeur, éventuellement à distance.

# 8. Conclusions

WinPitch LTL est un système ouvert et performant, bénéficiant de nombreuses fonctions auteurs pour l'élaboration de leçons d'exemples. Ces fonctions rendent l'adaptation au système de matériel déjà disponible particulièrement aisée. Ainsi des adaptations partielles de leçons de la méthode ASSIMIL ont été réalisées avec un minimum d'efforts pour l'anglais et le mandarin. Appliqué au mandarin, le logiciel utilise un codage couleur pour chacun des tons, ce qui s'avère très concluant pour l'apprentissage de cette langue. En particulier lorsque ce codage permet de se rendre compte que les tons théoriques tels que présentés dans des manuels ne sont pas toujours réalisés de la manière attendue, du fait de nombreux phénomènes de coarticulation et de sandhi.

Les points forts :

a. Système auteur ouvert, permettant l'élaboration d'exemples annotés indépendamment d'une théorie prosodique particulière.
b. Alignement automatique des segments de l'énoncé modèle sur celui de l'apprenant, avec surlignage et codage couleur des mouvements mélodiques importants.
c. Fonction de ralenti, assurant à l'apprenant une meilleure perception des mouvements mélodiques et leur relation avec la représentation graphique en courbes mélodiques.
d. Morphing prosodique, donnant à l'apprenant une réalisation selon le modèle mais utilisant sa propre voix.

Une version prochaine du logiciel devrait de plus donner un message texte des erreurs et différences des réalisations de l'apprenant, ainsi qu'une brève explication théorique.

# Références

ABBERTON, E. and FOURCIN, A. (1975). Visual feedback and the acquisition of intonation. *Foundations of Language Development, eds. E.H. Lenneberg and E. Lenneberg*. 157-165. New York: Academic Press, 1975.

BENALI I., CHIARELLI R., YOO H-Y., MARTIN Ph. (2006). Enseigner la prosodie avec le logiciel WinPitch LTL : possibilités technologiques et pédagogiques. N*ouvelles technologies et éducation en milieux formel et informel*, Casablanca, 2006.

BOT, K.D. (1980). The role of feedback and feedforward in the teaching of pronunciation − an overview. *System 8*: 35-47. 1983. "Visual feedback of intonation I: effectiveness and induced practice behavior." *Language and Speech 26 (4)*: 331-350, 1980.

CHAMPAGNE-MUZARD C., BOURDAGES J. S., (1998-1993). Le Point sur la phonétique. *CLE International*, 1998/1993.

GUIMBRETIERE E. (2000). Apprendre, Enseigner, Acquérir : La prosodie au cœur du débat. *Publications de l'université de Rouen*, 2000.

JAMES, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *IRAL 14 (3)*:227-243, 1979. "Intonation through visualization." *Current Issues in the Phonetic Sciences, eds. H. A. P. Holien*. 295-301. Amsterdam Studies in the Theory and History of Linguistic Science, IV, Amsterdam: John Benjamins.

LÉON, P.R. and MARTIN, Ph. (1972). Applied Linguistics and the Teaching of Intonation. *Modern Language Journal 56 (3)*: 139-144, 1972.

GERMAIN, A. et MARTIN, Ph. (2000). Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde. *www.alsic.org, 3, No 1*, 61-76, 2000.

JUN S-A. & FOUGERON, C. (2002). The Realizations of the Accentual Phrase in French Intonation, *Probus* 14, 147-172.

MARTIN, Ph. (1975). Analyse phonologique de la phrase française. *Linguistics*, (146) 35-68, Fév. 1975.

MARTIN, Ph. (1981). *Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne*. XIIème Journées d'Etude sur la Parole, Montréal, 223-232, mai 1981

MARTIN, Ph. (1982). Utilisation d'un visualiseur de mélodie en vue d'une didactique. *Options nouvelles en didactique du français langue étrangère.* Paris : Didier, pp. 181 – 186, 1982.

MARTIN, Ph. (2009). *Intonation du français*. Paris : A. Colin, 2009.

MARTIN, Ph. et DELAIS-ROUSSARIE, E. (2016) Comparison of three automatic speech segmentation systems, Proc. Interspeech 2016. San Francisco (submitted).

MOULINES, E. & CHARPENTIER, F. (1990). Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones. Speech Communication, vol. 9, nrs. 5/6, pp. 453-467.

STEVENS, V., SPURLING, S., LORITZ, D., KENNER, K., ESLING, J. and BRENNAN, M. (1986). New Ideas in Software Development for Linguistics and Language Learning. *CALICO Journal 4 (1)*: 15-26, 1986.

VARDANIAN, R. (1964). Teaching English intonation through oscilloscope displays. *Language Learning 14*: 109-118, 1964.

WELTENS, B. and BOT, K.D. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech 27 (1)*: 79-88, 1984.

WINPITCH LTL. (2015). http://www.winpitch.com/