



HAL
open science

Correction dynamique et adaptative de la justesse en voix chantée

Daniel Hernan Molina-Villota, Christophe d'Alessandro, Olivier Perrotin

► **To cite this version:**

Daniel Hernan Molina-Villota, Christophe d'Alessandro, Olivier Perrotin. Correction dynamique et adaptative de la justesse en voix chantée. CFA 2022 - 16ème Congrès Français d'Acoustique, Apr 2022, Marseille, France. hal-03673868

HAL Id: hal-03673868

<https://hal.science/hal-03673868v1>

Submitted on 25 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



16^{ème} Congrès Français d'Acoustique
11-15 Avril 2022, Marseille

Correction dynamique et adaptative de la justesse en voix chantée

Daniel Hernan Molina Villota^a, Christophe d'Alessandro^a, and Olivier Perrotin^b

^aSorbonne Université, CNRS, Institut ∂ 'Alembert, UMR 7109, LAM, Paris

^bINP, GIPSA-lab - Univ. Grenoble Alpes, 11 Rue des Mathématiques 38610 Gières



This study presents the application of Pitch Warping and Tuning (PWT) method to correct pitch from audio signals on real time. PWT use the dynamic pitch warping (DPW) method over f_o of an incoming audio signal. Through a pitch velocity condition, the correction is released and adapted to the constantly changing pitch values, in that way the curve of pitch is continuously adapted to the entry. At the end f_o a soft curve on real time is generated and the audio from incoming signal is forced to follow this curve. Several tests have been done using sinusoidal signals and singing voice audio on real time, these results show a significant difference with the traditional pitch correction systems. Our proposal shows a soft variation of pitch that is released over a pitch's velocity condition with an adaptive continuous modification

1 Introduction

Nous menons un projet d'étude systématique des effets audionumériques appliqués au style vocal. Dans ce cadre, il nous a semblé important de travailler sur les effets d'intonation. Le propos est de rendre ces effets dynamiques, sous le contrôle du chanteur, pour qu'il puisse changer au cours de la performance les paramètres des effets. De façon analogue au guitariste électrique, qui utilise depuis longtemps des effets audionumériques dynamiques, le chanteur pourrait ainsi changer le son de sa voix ou son style mélodique et s'emparer ainsi de nouveaux moyens d'expression.

La correction de justesse occupe une place de choix parmi les effets audio-numériques appliqués à la voix. Le succès considérable d'AutoTune de la société Antares (abrégée ATA dans cet article) ne se dément pas depuis plus de 25 ans. De façon inattendue, les artefacts sonores générés par des changements brutaux d'intonation ont très vite attiré l'attention des musiciens de façon positive. En effet, la qualité électronique particulière obtenue a été appréciée dans les musiques populaires, au point de devenir une marque de style musical, et d'être utilisée de façon quasiment systématique. L'effet change en effet la dynamique et l'identité de l'instrument vocal, permettant des transitions de notes beaucoup plus rapides que celles de la voix naturelle. Ces transitions vocales rapides donne un peu à la voix le caractère d'un instrument trous de jeu comme la flûte ou le saxophone, voire d'un instrument à clavier. Comme la rapidité des transitions viole les lois de la physique de la production vocale, le son prend une coloration électronique. ATA modifie l'intonation à l'aide d'un vocodeur, sans prendre en compte les variations naturelles du conduit vocal. Ces transitions rapides s'accompagnent alors de transitoires spectraux, qui sont devenus une signature acoustique de la méthode. Ainsi, ATA conçue au départ comme une méthode transparente du point de vue sonore, a acquis par ses artefacts le statut d'effet audionumérique bien identifié.

Dans cet article une nouvelle méthode pour la correction d'intonation vocale, Dynamic Pitch Warping, ou déformation mélodique dynamique (abrégée DPW dans cet article), qui peut fonctionner en temps-réel, est proposée. Cette méthode est l'adaptation à la voix d'une méthode de correction de trajectoire chironomique pour la synthèse vocale performative [3]. Après la présentation de la méthode DPW, les questions d'analyse de la fréquence de voisement et de modification par un vocodeur de l'intonation sont étudiées. Une évaluation comparative avec Autotune est ensuite

proposée. Les perspectives ouvertes pour le développement d'effets audionumériques vocaux dynamiques et contrôlés par le chanteurs sont évoquées.

2 Méthodes de correction mélodique

2.1 Correction mélodique ATA

ATA est une méthode audionumérique de correction mélodique, destinée au début, aux traitements de studio. Cette méthode a été développée par H. Hildebrand et brevetée [10] par Antares Audio Technologies en 1997 sous le nom d'Autotune.

ATA est aujourd'hui utilisé de façon massive dans l'industrie musicale. Lors de son apparition le principal défaut d'ATA a été la transition prononcée entre les notes. Mais ce défaut a été aussi la raison de son succès. L'utilisation comme correction transparente, en temps différé (studio), est presque systématique dans les productions commerciales [11]. L'exploitation des distorsions d'ATA a trouvé un "niche" dans des genres comme le rap, l'hyperpop et la musique électronique. ATA possède au départ une seul paramètre de contrôle : *retune speed* qui défini le temps de transition jusqu'à la note correcte visée. Le logiciel actuellement commercialisé possède un second paramètre de contrôle, *flex tune*, qui est l'intervalle en Cents MIDI dont la fréquence fondamentale peut s'écarter sans déclencher la correction. En plus de ces deux paramètres principaux, le logiciel offre plusieurs paramètres pour changer la qualité sonore. On peut par exemple choisir différentes échelles musicales ou changer la tonalité (*mineur, majeur*), conserver les variations subtiles de hauteur dans les notes longues et stables avec *humanize*, désactiver la modification des formants (*formants*) et modifier le modèle du conduit vocal, modifier la fréquence de référence pour l'échelle musical (*transpose, detune*), et ajouter du vibrato (réglable) aux notes tenues (*vibrato*). Pour l'analyse de fréquence fondamentale, plusieurs type de voix sont proposées, ce qui facilite les calculs en spécifiant l'ambitus moyen et en évitant ainsi les erreurs d'octave.

ATA n'est pas le seul correcteur de hauteur. D'autres dispositifs pour la modification de hauteur, avec des fonctionnalités similaires sont par exemple des dispositifs numériques comme metaTune, WavesTune et Ztx, et des dispositifs électroniques comme Model BOSS Vocal Performer, Boss VE-20, TC Helicon, Tascam TA-1VP et Roland VT-4. Mais les méthodes de correction de hauteur ne s'appliquent pas qu'aux effets vocaux. On trouve des

méthodes de corrections également dans les instruments musicaux numériques. Dans ce cas, la correction améliore la précision de la capture du geste et la facilité de jeu musical pour les utilisateurs. Parmi ces instruments musicaux on peut citer [15] Continuum Fingerboard2 [16], Seaboard3 [17], Garageband, TouchKeys et Cantor Digitalis. Le Cantor Digitalis [2] comprend une méthode de correction de hauteur du tracé obtenue par interface tactile. Cette méthode [3] est basée sur une fonction de déformation dynamique dénommé *Dynamic Pitch Warping*.

Dans cet article la méthode DPW est mise en œuvre non pour corriger une trajectoire graphique, mais la courbe de hauteur d'un signal vocal d'entrée.

2.2 Méthode de correction DWP

La fonction proposé par DPW est nommé *fonction élastique* [4]. Cette méthode possède trois degrés de liberté, l'intervalle de détection, le temps critique, et le temps de transition. L'intervalle de détection et temps critiques se combinent pour calculer la condition indirecte de vitesse mélodique, qui détermine l'activation de la correction. La vitesse mélodique permet de définir un seuil de temps critique de détection dans un intervalle de détection. Cela permet d'inférer si la note est assez stable, intentionnelle, et donc mérite d'être corrigée. Si la variation mélodique est trop rapide, la note n'est pas stable et il ne faut pas déclencher de correction. L'avantage des trois degrés de liberté est de permettre la préservation des variations mélodiques expressives, comme le vibrato, en ne corrigeant pas systématiquement la hauteur.

Dans la version initiale de DPW, les problèmes pratiques de détection et de modification de la fréquence fondamentale ne se posent pas : l'entrée donne exactement la fréquence fondamentale mesurée sur le capteur, et ensuite le synthétiseur vocal la calcule directement.

3 Système de correction mélodique

3.1 Architecture

Le système de correction mélodique utilisant la fonction DPW travaille de la façon suivante. L'interprète chante; on capture le signal sonore et on calcule la fréquence fondamentale. Si la hauteur reste stable dedans un intervalle dans le voisinage d'une note (intervalle de détection, I) pendant un seuil temporel (temps critique, t_c), le système déclenche la correction. Cette condition peut être traduite comme une condition de vitesse, cela signifie que les changements rapides de notes tels que les trajectoires naturelles de la voix ne vont pas être corrigés et que seulement les notes stables vont être corrigées. La correction est faite de manière douce pendant une période de transition et de façon adaptative, ainsi les notes suivantes seront aussi accordées grâce à la fonction élastique.

La figure 1 montre la fonction de notes étendues utilisée dans certaines méthodes de correction et la fonction élastique développée par Perrotin et D'Alessandro. Le zéro représente

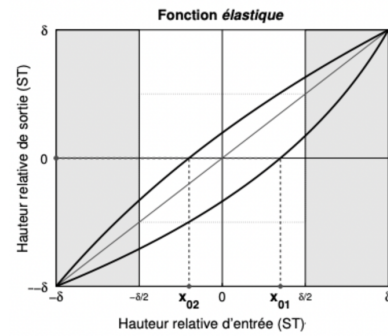


FIGURE 1 – Méthode pour la correction de hauteur [3]

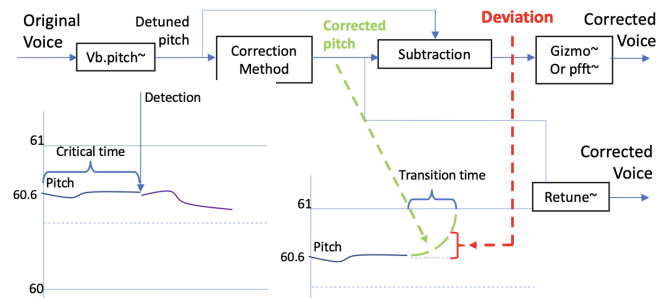


FIGURE 2 – Schéma du DPW

le point de l'échelle musicale de référence. $\delta = 1$ représente une différence d'un degré (un demi-ton le plus souvent) avec la note suivante et précédente. On utilise les conditions limites $y(\pm\delta) = \pm\delta$. et le fonction élastique suit :

$$g(y) = Ae^{\gamma(y+B)+C} \quad (1)$$

Pour calculer la solution, C est remplacée dans $g(y)$ et on efface ainsi la dépendance en B. Avec $y(\pm\delta) = \pm\delta$ on obtienne :

$$y_E(x) = \begin{cases} \frac{1}{\gamma} [\log[(e^{2\gamma} - 1)(\frac{x}{\delta} + 1)^{\frac{1}{2}} + 1]] - \delta, & \text{if } \gamma \neq 0 \\ x, & \text{if } \gamma = 0 \end{cases} \quad (2)$$

x est la hauteur d'entrée et y_E est la hauteur de sortie. Il ne doit pas y avoir correction pour les valeurs entières, donc $y_E(x_o, \gamma_o) = 0$, et on peut obtenir la courbure :

$$\gamma = \frac{1}{\delta} \log\left(\frac{\delta - x_o}{\delta + x_o}\right) \quad (3)$$

Pour δ donné, la valeur de γ est déduite a partir de. La sortie est accordée seulement si la hauteur d'entrée est stable. Quand l'entrée prend les valeurs $x_{integer} \pm 1$, on a $y(\pm\delta) = \pm\delta$. Quand la correction est désactivée, γ revient graduellement à 0, et peut s'activer encore dès que la condition de vitesse est remplie de nouveau. Un schéma du fonctionnement est donné figure 2.

3.2 Adaptation à la voix

3.2.1 Analyse de la fréquence fondamentale d'entrée

Le premier pas pour adapter la méthode à la voix, est d'obtenir la courbe *fzero* d'entrée. Pour cela les méthodes suivantes ont été testées : *yin* [14], *vb.pitch* (V. Bohm), *sigmund and fzero* [8]. Les meilleurs résultats sont obtenus avec l'objet *fzero*, avec la configuration suivante : `@onsetamp 0.0001 @onsetpitch 0.001`, les seuils de détection d'amplitude et de hauteur. La sortie (*fzero*) est ensuite dirigée vers le patch de correction DPW qu'on a créé. Les paramètres contrôlables dans ce patch sont les suivants :

- *Fzero*,
- Temps de détection t_c ,
- Temps de transition
- Liste de notes activées/désactivées

Une fois obtenue la courbe de *fzero*, elle est passée dans le filtre des notes activées. Dans le patch assigné à ce filtre, on localise *fzero* dans une octave. Dans l'octave on connaît les notes activées ou désactivées. Si *fzero* est dans un intervalle activé, il passe à la sortie du filtre. Sinon on calcule la note activée la plus proche et on obtient la valeur extrême de l'intervalle en restant dans le voisinage d'une note activée. Quand la fréquence évolue entre régions activées en passant par des régions désactivées, on utilise les valeurs extrêmes des régions désactivées, cela permet d'utiliser uniquement les régions activées et de contrôler la vitesse de transition entre les zones interdites. Ce temps de transition entre notes désactivées peut être considéré comme un 4^{ème} degré de liberté. L'utilisation d'un objet *gate* permet de bloquer le passage de *fzero* quand on est sur les zones interdites. Dans les régions activées *fzero* passe directement à la sortie.

Le filtre de notes activées et désactivées est important car il permet de contrôler l'échelle mélodique de l'interprète. La différence avec ATA provient des degrés de liberté. ATA propose seulement le contrôle du temps de transition, identique pour les intervalles activés et désactivés, alors qu'avec DPW on a le temps de détection, le temps de transition, les intervalles de détection et le temps de transition entre notes interdites.

3.2.2 Correction de l'intonation par DPW

La condition pour déclencher la correction est basée intrinsèquement sur l'idée de stabilité. Seulement les notes d'une durée supérieure à t_c et comprises dans un intervalle I déclencheront la correction de l'algorithme. Toutes les autres entrées sont interprétées comme des trajectoires de hauteur naturelle et ne seront pas corrigées. Comme noté précédemment, la fonction élastique fait que la sortie et l'entrée convergent dans les extrêmes vers les valeurs entières. Si aucune correction n'est nécessaire (la condition de déclenchement n'est pas remplie), alors l'entrée passe directement à la sortie.

Pour la mise en œuvre de la méthode DPW, nous avons besoin des paramètres des 3 degrés de liberté plus le temps de transition pour les notes désactivées. Avec l'intervalle de détection, nous divisons l'échelle musicale en micro-tons de 0,1 demi-tons (ST). Cela signifie que l'intervalle $(\delta, -\delta)$ est découpé en tranches de 0.1 ST. Si la hauteur d'entrée reste dans l'une des sections de micro-tons pendant le temps de détection, la correction est déclenchée pendant le temps de transition. Le résultat est une courbe de hauteur déformée où les notes suffisamment stables sont corrigées. L'autre degré de liberté est le temps transition (t_t) est comparable à *retune speed*. On peut considérer aussi un quatrième degré de liberté qui est le temps de transition entre notes désactivées.

3.2.3 Resynthèse de l'intonation modifiée

Une fois la courbe mélodique corrigée calculée, il faut utiliser un vocodeur pour la resynthèse du signal correspondant. Le premier vocodeur a été breveté par Homer Dudley aux Bell Labs en 1935 pour reconstruire une voix à l'aide d'un algorithme d'analyse-synthèse, il avait été créé pour réduire la bande passante d'un signal vocal $v(t)$ pour les télécommunications, mais finalement appliqué et la musique et aux bandes son de films [11]. Aujourd'hui de nombreux types de vocodeurs existent : channel vocoder qui utilise la somme pondérée des filtres passe-bande, le LPC qui utilise un filtre IIR, et le cepstrum qui utilise une convolution circulaire d'un filtre FIR. [13]

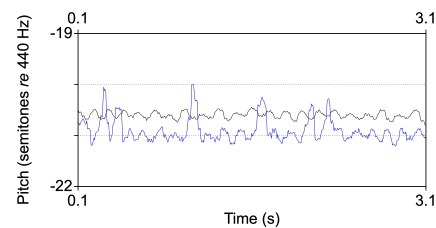


FIGURE 3 – Courbe intonative avec vibrato (noir) et sa version corrigée avec DPW (bleu).

Les vocodeurs testés pour notre système utilisent les objets suivants : *pitchshift*, *freqshift*, *psych*, *supervp.trans*, *fbinshift*, *transpose*, *hilbert*, *gizmo* et *retune*. Tous sauf *retune* reçoivent en entrée la valeur en midicents à transposer. On ne peut généralement pas accéder à la détection de seuil de cette entrée, et les objets ne sont pas assez réactifs. Avec *supervp* et *gizmo*, la réponse est meilleure mais n'est pas assez précise pour notre objectif. Le meilleur résultat a été obtenu avec *retune* un objet Max MSP basé sur *zynaptiq ztx* (Precision Time Stretching and Pitch Shifting) [12] une méthode brevetée, comme ATA. L'objet *retune* est un vocodeur qui sert de base pour les vocodeurs natifs d'Ableton tels que *mono vocoder*, *poly vocoder* et *autotuna*. Cet objet peut recevoir une entrée (signal) pour forcer le pas de sortie souhaité. Nous utilisons cette possibilité pour déformer la hauteur de sortie avec notre correction. Actuellement notre système ne dispose pas

des fonctions supplémentaire d'ATA telles que la sélection du type de voix, vibrato et humanisation.

4 Exemples de corrections et évaluation comparative

L'évaluation de la méthode est conduite en deux temps. Dans un premier temps, il s'agit d'analyser le comportement de la méthode, et ainsi de vérifier si elle répond à nos attentes. Dans un deuxième temps, la nouvelle méthodes est comparée à ATA dans divers cas de correction, afin de mettre en évidence les différences entre les deux méthodes et les avantages possibles de la nouvelle méthode.

Pour la phase d'évaluation initiale, les sons de référence sont générés par un synthétiseur vocal performatif, le Cantor Digitalis [2]. Ce système permet un contrôle chironomique très précis de l'intonation, et donc d'obtenir des sons vocaux de synthèse avec davantage de précision et de justesse que la voix naturelle [1]. L'objectif est de vérifier la réactivité de la méthode et son comportement pour des situations de transitions de notes et d'ornement. Les cas typiques du vibrato, du glissando et du portamento sont étudiés.

4.1 Vibrato

Le vibrato est un ornement musical courant, qui est perçu de façon complexe [9]. C'est une modulation de fréquence (mais aussi d'intensité et de timbre) d'une amplitude de l'ordre de quelques dizaines de Cents, et avec des fréquences inférieures de 4 à 6 Hz environ. Le vibrato ajoute de l'expression, de l'emphase, de la variété de timbre, renforce la présence des chanteurs [7]. La figure 3 montre une correction de hauteur qui préserve le vibrato. Pour un exemple similaire, dans le cas d'ATA sans les nouvelles caractéristiques, le vibrato est entièrement supprimé, et la hauteur corrigée saute d'un demi-ton à l'autre, en fonction de la vitesse de correction. Ici, le vibrato est assez rapide, et son amplitude de l'ordre de 25 Cents en valeur absolue. Avec le DPW, l'intonation moyenne est correctement modifiée, mais le vibrato est bien conservé, ce qui est un des buts de la méthode. Cependant, lorsque l'intonation est trop proche de la note (demi-ton) supérieure, la correction "accroche" cette note. Pour préserver le vibrato, il faut donc soigneusement régler les paramètres de la méthode. L'exploitation musicale de cette source d'artefact reste à explorer.

4.2 Glissando

Les glissandos, même s'ils sont rarement employés musicalement, sont très utiles pour analyser le comportement des méthodes de correction d'intonation. Les glissandos sont réalisés avec le Cantor Digitalis, sur une octave, avec une pente plus ou moins rapides. L'objectif du test est de vérifier si le système réagit correctement. La figure 4 montre les résultats pour différents réglages des paramètres de correction et différentes pentes de glissando. Pour l'image du haut, la pente est de 3 demi-tons/s. La correction est conforme aux attentes, avec la transformation du glissando

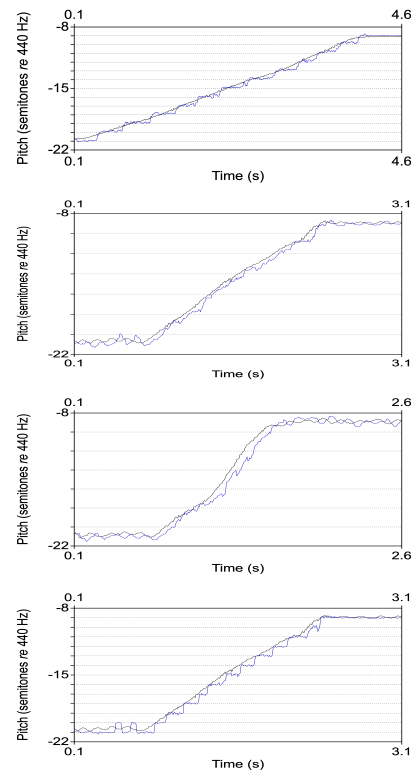


FIGURE 4 – De haut en le bas : Glissando sans vibrato de C3 à C4 pendant 4000ms. Glissando lent avec vibrato de C3 à C4. Glissando rapide avec vibrato de C3 à C4. Glissando lent avec vibrato de C3 à C4 avec une correction instantanée

en escalier mélodique, qui suit l'échelle des demi-tons, tout en préservant par ailleurs les micro-variations mélodiques. Notre correction apporte donc une certaine variabilité au son.

Des variations de vitesse du glissando, en présence de vibrato, sont analysées sur les deux images centrales. Dans ces exemples, la vitesse de déclenchement de la correction en fonction de la vitesse de variation mélodique permet de préserver la pente du glissando, tout en corrigeant le début et la fin des notes. Il est important de mentionner que la correction mélodique n'est pas immédiate ici. Le but de ces deux exemples est de vérifier que le glissando corrigé suit assez bien la forme initiale, mais avec une correction dans les notes extrêmes du son.

L'image du bas de la Figure 4 montre l'effet d'une correction instantanée de hauteur (c'est-à-dire avec un temps de détection et un temps de détection nuls). Le résultat est un escalier mélodique, avec des marches quasiment plates et des contremarches quasiment verticales. Ce résultat est très similaire à celui attendu avec ATA pour un temps de transition nul. La partie plate initiale du son est corrigée vers la note la plus proche, ensuite l'allure du glissando est conservée, mais en escalier, puis, la partie finale du son est corrigée. Dans cet exemple, il y a moins de variation de hauteur à l'intérieur d'une note stable, à cause du temps de détection utilisé qui est de 0.

4.3 Portamento

La figure 5 montre l'analyse dans le cas du portamento, ou transition "portée" entre deux notes.

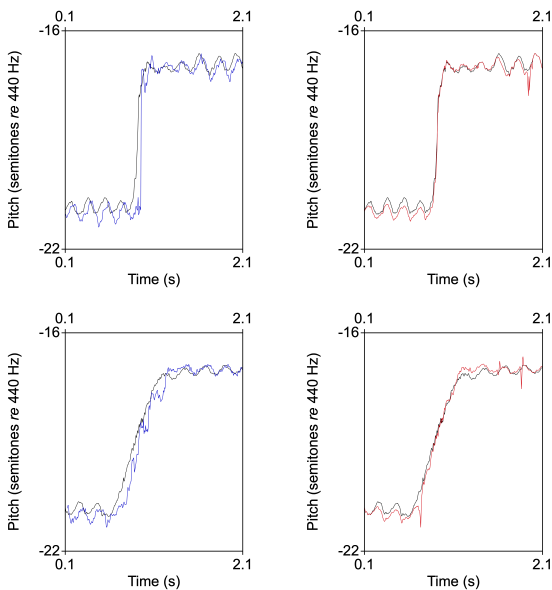


FIGURE 5 – En haut : DPW (15ms de détection et 30ms de transition) (bleu) et ATA (45ms)(rouge) pour un portamento rapide. En bas : DPW (15ms de détection et 30ms de transition) (bleu) and ATA (45ms)(rouge) pour un portamento plus lent.

Les résultats sont donnés sur la figure 5. La méthode donne une bonne correction, meilleure pour le portamento plus lent. Dans les deux cas la forme générale de la transition est conservée, et il y a une correction moyenne de hauteur qui préserve le vibrato pour les notes sautées avant et après la transition. Les images du bas montrent le même type de correction avec Autotune.

Nous pouvons observer quelques différences dues à un meilleur contrôle du grain de correction avec notre méthode. Pour une correction avec une constante de 45ms ATA ne corrige presque pas, alors que notre méthode apporte quelques modifications. Elle est plus rapide dans la correction qu'ATA. Notre méthode permet un meilleur contrôle du grain de correction comme on le voit sur la figure.

La figure 6 montre l'effet des variations de paramètres dans notre méthode. Un glissando est corrigé selon trois conditions différentes (temps de détection 7 ms, temps de transition 15ms, temps de détection 15 ms, temps de transition 15ms, temps de détection 30 ms, temps de transition 60ms) en magenta, bleu et marron respectivement, l'échantillon original est en noir. Comme nous pouvons le voir (et l'entendre) la modification devient moins audible, plus douce pour des temps de détection et de transition plus longs.

L'échelle mélodique de correction peut être spécifiée à volonté. Dans ce cadre, nous avons fait un test avec une échelle mélodique chromatique pour laquelle certaines

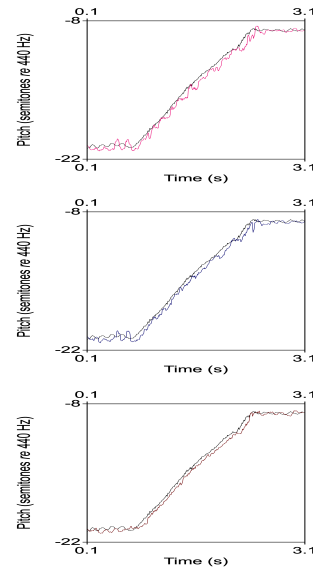


FIGURE 6 – Correction par DPW avec 3 réglages différents.

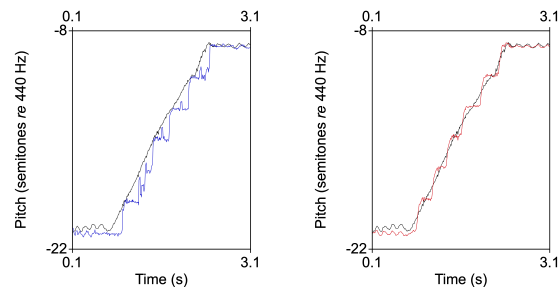


FIGURE 7 – Test avec la moitié des notes désactivées

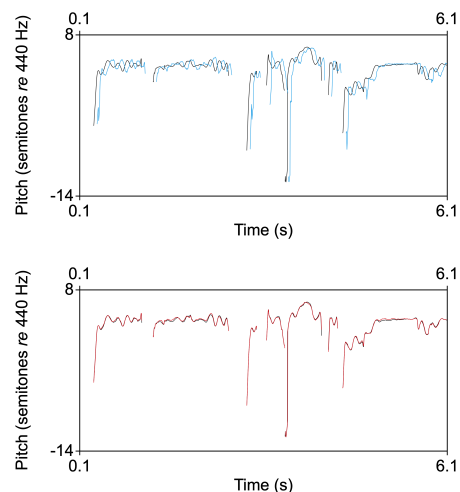


FIGURE 8 – Test avec un son réel corrigé avec DPW (bleu) et Antares (rouge)

notes sont interdites ou désactivées (notes désactivées = 1, 3, 5, 7, 9, 11; notes activées = 0, 2, 4, 6, 8, 10) avec un temps de détection de 7 ms et un temps de transition de

15ms. Le résultat est porté sur la partie gauche de la figure 7. La partie droite montre le résultat d'ATA pour un temps de 21 ms. Cette dernière correction semble de meilleure qualité.

Finalement une chanson est analysée avec notre méthode et ATA dans la figure 8. La correction est plus claire avec notre méthode pour une configuration de 45 ms de durée (15 ms pour le temps de déclenchement et 30 ms pour le temps de transition). Il faut remarquer que notre méthode n'est actuellement pas plus rapide qu'ATA, mais ne pose pas de problèmes de rendu sonore.

5 Discussion

Le premier objectif de cette recherche était l'application à la voix chantée, et en temps réel d'un nouvel algorithme de correction de hauteur, qui avait été développé pour la correction de cibles sur une tablette graphique. Nous avons ainsi montré que l'algorithme DPW, en lui adjoignant une détection de fréquence fondamentale en entrée et le vocodeur ZTX en sortie peut s'appliquer à la voix en temps réel. Il est ainsi possible d'ajouter des degrés de liberté pour le contrôle de l'effet. De plus le logiciel est ouvert, et donne des résultats sonores comparables à ceux d'ATA pour le temps réel. Les paramètres de contrôle permettent de ne corriger que les changements de note sur l'échelle et non les ornements mélodiques entre les notes. Le vocodeur lui-même semble introduire des imprécisions, et il sera nécessaire de travailler encore cet aspect. Ces imprécisions qui apparaissent visuellement sur les tracés semblent avoir peu d'effet auditif. Le fait d'avoir des contrôles différents introduit des artefacts différents, dont l'exploitation musicale doit être étudiée.

L'utilisation d'un contrôle dynamique de l'effet de correction est un de nos objectifs. Nous avons démontré que le système fonctionne de manière robuste. Notons que le système est ouvert, via MAX MSP. Ainsi la méthode peut s'appliquer dans divers systèmes ou modules d'effet.

L'utilisation de nouveaux paramètres permet de nouvelles possibilités pour la création et la performance musicale. Pour la suite du travail, une évaluation perceptive formelle doit être menée. L'étude de différents vocodeurs et des artefacts qui en résultent doit également être conduite. En effet les variations de timbre résultant des modifications des transitions naturelles sont de la plus haute importance pour la perception de l'effet sonore et pour son utilisation artistique.

Remerciements

Cette recherche a bénéficié d'un financement à travers les projets de l'Agence Nationale de la Recherche ARS : Analysis and Transformation of Singing Style (ANR-19-CE38-0001) et Gepeto : GESture and PEdagogogy of inTOnation (ANR-19-CE28-0018).

Références

- [1] C. d'Alessandro, L. Feugère, S. Le Beux, O. Perrotin, A. Rilliard, Drawing melodies : Evaluation of chironomic singing synthesis, *JASA* 135 (6) (2014) 3601–3612.
- [2] L. Feugère, C. d'Alessandro, B. Doval, O. Perrotin, Cantor digitalis : chironomic parametric synthesis of singing, *EURASIP Journal on Audio, Speech, and Music Processing* 2017 (1) 2.
- [3] O. Perrotin, C. D'Alessandro, Target Acquisition vs Expressive Motion : Dynamic Pitch Warping for Intonation Correction, *ACM Transactions on Computer Human Interaction* 23, 1-21 (2016).
- [4] O. Perrotin, Chanter avec les mains : interfaces chironomiques pour les instruments de musique numériques *Thèse de doctorat Informatique, U.Paris Saclay, Ch4*, 93-123 (2015).
- [5] V. Verfaille, M. Wanderley, P. Depalle, Mapping Strategies for Gesturaland Adaptive Control of Digital Audio Effects, *Journal of New Music Research* 35-1, 71-93 (2006).
- [6] V. Verfaille, C. Guastavino, C. Traube, An interdisciplinary approach to audio effect classification, *Proceedings 9th International Conference on DAFX*, 106–113 (2006)
- [7] L. Regnier, PhD Thesis - Localization, Characterization and Recognition of Singing Voices, *Université Pierre et Marie Curie - Paris VI*, Paris (2012).
- [8] M. Zbyszynski, D. Zicarelli, R. Collecchia, fzero Fundamental estimation for Max 6 *International Computer Music Conference*, Perth, Australia (2013).
- [9] C. D'Alessandro, M. Castellengo, The pitch of short-duration vibrato tones, *Journal of the Acoustical Society of America*, 93(3) 1617-1630(1994)
- [10] H. Hildebrand, Autotune Antares Patent, Pitch detection and intonation correction apparatus and method, *Auburn Audio Technologies, CA*, (1998)
- [11] C. Vincent, De l'antipop à l'Autotune, *La voix chantée, N. Henrich, Boeck solal Ch 8* 123-137(2014)
- [12] S. Bernsee, D. Gökdag, ZTX Patent - Frequency transform extension methods to solve characteristics in the space-time domain, *Zynaptiq GmbH*, (2015)
- [13] U. Zölzer, DAFX : Digital Audio Effects, *Wiley*, 315 (2002)
- [14] A. de Cheveigné, H. YIN, A fundamental frequency estimator for speech and music, *J. Acoust Soc Am*, 1917-1930 (2002)
- [15] Vincent Goudard, Hugues Genevois et Lionel Feugre : On the playing of monodic pitch in digital music instruments. In Anastasia Georgaki et Giorgos Kouroupetoglou, diteurs : Proceedings of the International Computer Music Conference (ICMC), pages 1418–1425, Athens, Greece, September 2014. National and Kapodistrian University of Athens.
- [16] Lippold Haken : Position correction for an electronic musical instrument, novembre 17 2009. US Patent 7,619,156.
- [17] Roland Lamb et Andrew N. Robertson : Seabord : a new piano keyboardrelated interface combining discrete and continuous control. In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), NIME '11, pages 503–506, Oslo, Norway, May 30 - June 1 2011.
- [18] IRCAM, SuperVP-TRaX, <http://forumnet.ircam.fr/373.html?L=1>.