



Étude exploratoire des stratégies de production du ton 3 en chinois mandarin

Yizhi Huang¹, Véronique Delvaux^{1,2}, Kathy Huet¹, Myriam Piccaluga¹, Guoxian Zhang¹,
Bernard Harmegnies¹

(1) Institut de Recherche en Sciences et Technologies du Langage, UMONS, Belgique

(2) FNRS, Belgique

yizhi.huang@umons.ac.be

RESUME

Malgré l'absence de contraste linguistique au niveau du mode de phonation en chinois mandarin, des études antérieures ont documenté, lors de la production des tons, de fréquentes occurrences de phonation non modale (type "creaky voice"/"vocal fry"), en particulier avec le troisième et le quatrième ton. La f₀ étant traditionnellement considérée comme l'élément déterminant dans la production des tons, leurs autres corrélats acoustiques/phonétiques demeurent inexploités dans ces études antérieures. Sur la base d'un corpus constitué de mots et de pseudo-mots monosyllabiques de structure CV (où V est [a]) (tons T1, T2, T3, T4), et C [m]) produits isolément par 10 locuteurs natifs du chinois mandarin, la présente étude vise à décrire la distribution des occurrences de phonation non modale en fonction des paramètres structurant le corpus, et à mieux caractériser les stratégies de production du T3 en particulier, à partir de mesures acoustiques diversifiées.

ABSTRACT

Preliminary study on strategies of native T3 production in Mandarin Chinese

Non-modal phonation is present in Mandarin Chinese speech production but marks no linguistic contrast. In previous studies, T3&T4 are reported to sometimes be produced with creaky voice/vocal fry. The pitch is normally considered as the primary property in production, but other acoustic/phonetic correlates of the Mandarin Chinese lexical tones remain not fully understood. Basing on a corpus consisted of monosyllabic words and non-words produced in citation form by 10 native speakers of Mandarin Chinese, the present study aims at describing the occurrences of non-modal phonation in Mandarin Chinese, particularly in T3 production, and the characteristics of the strategies which are implemented in tone production with various acoustic measurements.

MOTS-CLES : tons; chinois mandarin; voix craquée; phonation

KEYWORDS : tones; mandarin; creaky voice; phonation

1 Introduction

Le chinois mandarin est doté d'un système de contraste tonal à cinq tons, dont un ton statique (T1), trois tons dynamiques (T2, T3 T4) et un ton neutre. Une méthode de transcription très adoptée par

les chercheurs du domaine qu'est la numérotation de *Chao* (Chao, 1930) référencée comme Chao digit par certains phonologues (Yip.M., 2003 ;Duanmu. S, 2007, 226) où le registre et le contour de pitch sont représentés sur une échelle à cinq niveaux numérotés de 1 à 5, du registre le plus bas vers le plus haut. Les quatre tons contrastifs du mandarin peuvent ainsi être transcrits comme 55 (T1), 35 (T2), 214 (T3), 51 (T4), ce qui permet de marquer le registre et la dynamique de pitch d'un ton. Traditionnellement, la fréquence fondamentale (f0 moyenne, évolution temporelle de la f0) est considérée comme le principal corrélat acoustique des tons. Quelques études ont évoqué d'autres indices phonétiques/acoustiques qui pourraient jouer un rôle dans la production et la perception des tons - parmi lesquels le voisement de la consonne précédente, la durée syllabique et la qualité vocale/le mode phonatoire - mais ces éléments demeurent globalement peu exploités dans la littérature décrivant la production et, plus encore, la perception des tons (Kuang, 2017).

En ce qui concerne le rapport entre mode phonatoire et tons, des études antérieures ont suggéré une covariation entre la phonation non-modale de type *creaky* et le pitch en chinois mandarin, le T3 et le T4 étant parfois réalisés avec une qualité de voix dite « *creaky voice* » (Chao, 1956 ; Davidson, 1991; Belotel-Grenié et Grenié, 1994; Kuang, 2017). Rappelons qu'en chinois mandarin, la phonation non-modale de type « *creaky voice* » ne supporte pas de contraste linguistique (Moisik et al., 2014). Il s'agirait plutôt ici de variation allophonique typiquement associée à une fréquence fondamentale très basse (Kuang, 2017). Le chinois mandarin n'est pas la seule langue présentant ce type de covariation. En vietnamien du nord (Brunelle, 2009) et en cantonais (Yu et Lam, 2014), ce mode phonatoire particulier est employé de manière allophonique avec les tons. Yu et Lam (2014) ont étudié le rôle du « *creaky voice* » dans la perception des tons en cantonais, et ont montré une amélioration de 20% du taux d'identification du T4 (un ton avec un registre bas) lorsque les stimuli étaient accompagnés du mode phonatoire « *creaky* » par rapport aux stimuli produits en phonation modale. Une étude récente (Yang, 2011) a également montré que l'emploi de « *creaky voice* » est particulièrement utile pour discriminer le T2(35) et T3(214), la paire tonale la plus confondue en perception et en production chez les natifs du mandarin, car le T3 (214) est souvent produit avec un profil tonal ressemblant à celui du T2(35) dans le contexte sandhi T3T3 ou T2T3. Puisque seul le contexte monosyllabique est présent, la particularité du ton neutre (il s'agit d'une absence de ton pour les syllabes non accentuées) n'est pas examinée ici.

L'objectif de la présente étude exploratoire est de décrire acoustiquement la production du T3 à partir des productions de 10 natifs du chinois mandarin, avec une attention particulière portée à la variabilité inter-individuelle et aux potentielles stratégies impliquant l'emploi d'indices acoustiques complémentaires à l'évolution de la fréquence fondamentale. Premièrement, l'évolution de la f0 sera étudiée et comparée à l'attente phonologique (T3=214), tant au point de vue du groupe que des profils individuels. Ensuite, on s'intéressera à la présence de très basses fréquences et à leur éventuel corollaire, le mode phonatoire de type "*creaky voice*". Enfin, on recherchera la présence de corrélats acoustiques autres que ceux associés à la f0 lors de la production du T3, principalement les variations de timbre et de durée vocalique. L'orientation de cette étude exploratoire est justifiée par notre interrogation sur l'importance relative des diverses stratégies de réalisation des tons du chinois mandarin, et ce dans une perspective ultérieure d'implémentation des observations réalisées auprès des locuteurs natifs dans le contexte de développements en didactique du Chinois langue étrangère à l'intention de natifs du français. Les données recueillies dans la présente étude seront par ailleurs réutilisées dans de futures expériences en perception des tons lexicaux par des apprenants francophones.

2 Méthodologie

2.1 Corpus et participants

Le corpus est constitué de mots et de pseudo-mots monosyllabiques de structure CV où C est [m] et V est [a], portant les 4 tons lexicaux: T1, T2, T3, T4. Les items du corpus ont été répétés 4 fois de façon isolée par 10 natifs du chinois mandarin, 5 hommes et 5 femmes âgés de 18 à 36 ans, dont la plupart venant des régions où les dialectes de la famille mandarin (Pékin 1, Hebei 1, Henan 1, Shandong 2, Heilongjiang 1, Nanjing 2, Hubei 1) sont parlés, y compris le mandarin pékinois. Seule personne parlant un dialecte (Wu) hors de la famille des dialectes mandarin est l'auteur de l'article, ayant obtenu le certificat national de Putonghua. Les sessions d'enregistrement se sont déroulées dans une chambre sourde, les participants devant produire des stimuli en idéogrammes et pinyin affichées sur un écran d'ordinateur en face d'eux. La prononciation et le sens des idéogrammes moins connu sont préalablement expliqués à chaque participant. Les enregistrements ont été effectués à l'aide d'un enregistreur digital zoom H5.

2.2 Mesures

Les mesures ont été effectuées manuellement à l'aide du logiciel Praat. La fréquence fondamentale (f_0 , en Hz) et les 4 premiers formants (F1, F2, F3, F4, en Hz) ont été mesurés en début et fin des syllabes portant le ton 3, (les données recueillies sur les trois autres tons sont utilisées uniquement en comparaison avec la durée moyenne du T3 dans le présent article). La durée de la syllabe ainsi que la durée de la consonne initiale ont également été mesurées. Enfin, on a codé chaque occurrence de T3 pour la présence (1) ou l'absence (0) de "creaky voice" sur base de critères auditifs (jugement expert par le premier auteur, natif du mandarin) et visuels: présence d'irrégularités au niveau des pulses glottaux, détection d'un pitch discontinu par Praat, changement important d'amplitude et/ou de fréquence entre cycles consécutifs visibles sur le signal de parole et/ou le spectrogramme. Notons qu'en cas de rupture dans la détection de f_0 par l'algorithme de pitch de Praat, bien que la périodicité des pulses glottiques demeure, la f_0 a été évaluée directement à partir du signal de parole sur la période la plus longue, via la formule: $f_0(Hz)=1/T(ms)$.

3 Résultats

3.1 Fréquences fondamentales moyennes

Comme le montre la figure 1, les fréquences fondamentales du T3(214) observées sont, en moyenne, plus élevées chez les femmes que chez les hommes, comme on pouvait évidemment s'y attendre, étant donné les différences de conformation physique. On observe, de manière plus intéressante, que les valeurs calculées présentent des tendances cohérentes avec l'attente phonologique puisque,

tant chez les locuteurs féminins que chez les locuteurs masculins, apparaît la forme en V, qui trahit une baisse sensible du pitch au centre de la syllabe.

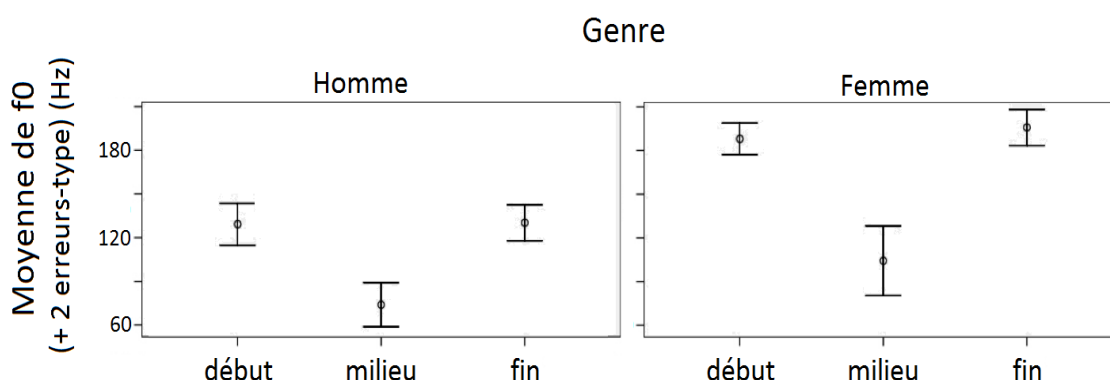


Figure 1 : Moyenne et intervalle de confiance (moyenne \pm 2 erreurs types) des valeurs de fréquence fondamentale (en Hz) relevées en début, milieu et fin de syllabe portant le T3 pour les hommes (en haut) et les femmes (en bas).

Dans chacun des deux groupes de genre, le fondamental moyen de fin de syllabe est, par ailleurs, supérieur à celui du début (hommes : 130,20 Hz > 129,25 Hz; femmes : 195,85 Hz > 188,05 Hz) ; cette différence est cependant extrêmement ténue.

En outre, si les valeurs de début et de fin de syllabe sont certes légèrement élevées pour chacun des groupes de genre mais cependant assez banales (de l'ordre de 130 Hz chez les hommes et de 190 Hz chez les femmes), elles sont spécialement basses en milieu de syllabe (74 Hz dans le groupe masculin et 104 Hz dans le groupe féminin). On constate aussi que les variances sont maximales au centre de la syllabe (femmes : $\sigma = 53$ Hz > 24 Hz et 27 Hz; hommes : $\sigma = 33$ Hz > 32 Hz et 27 Hz).

3.2 Fréquences fondamentales: variabilité interindividuelle

La figure 2 permet un regard plus précis sur la variabilité interindividuelle caractérisant les valeurs de fréquence fondamentale. Dans chaque groupe de genre, elle présente, pour chaque sujet, ses valeurs de f0 mesurées en début et fin de syllabe (reliées par un segment de droite), et sa valeur mesurée en milieu de syllabe (représentée par un cercle), le tout rassemblé en un rectangle vertical grisé. La figure 2 permet de prendre la pleine mesure de l'importante variabilité interindividuelle présente tant dans le groupe masculin que dans le groupe féminin, les sujets se dispersant largement sur l'axe des fréquences fondamentales.

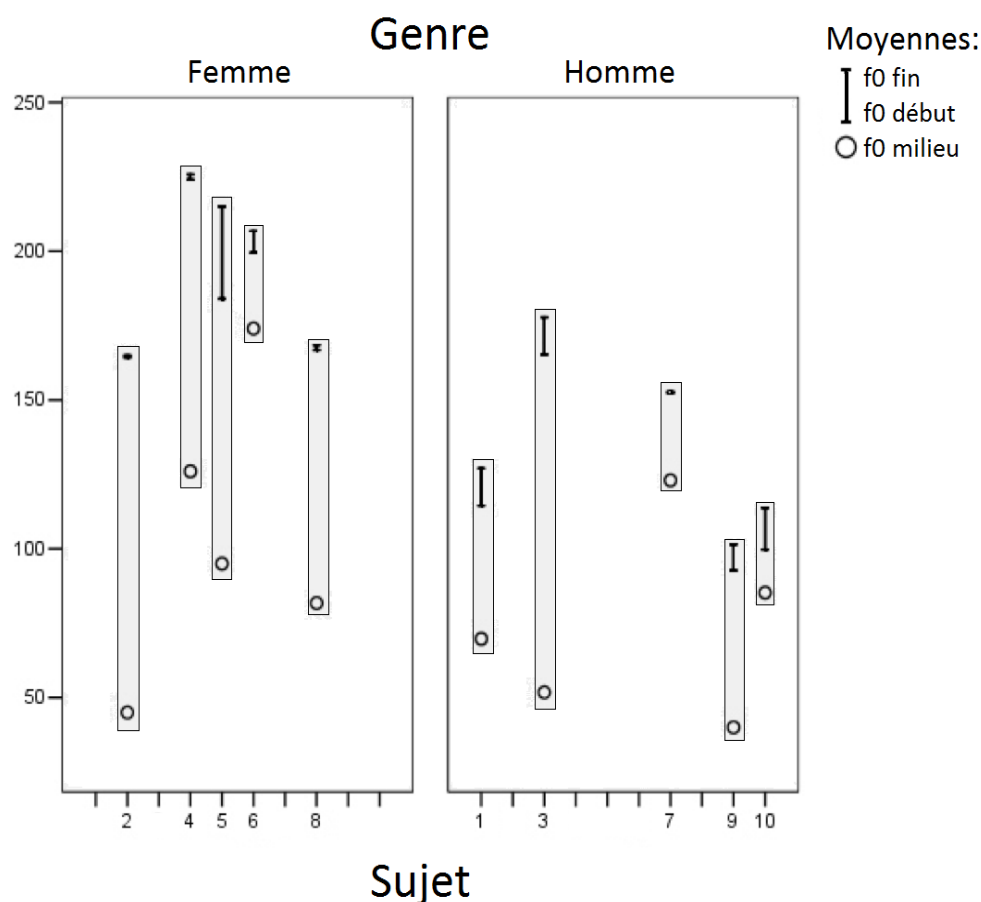


Figure 2 : Fréquence fondamentale (en Hz) en début et fin de syllabe (valeurs reliées par un segment de droite) ainsi qu'en milieu de syllabe (valeur figurée par un cercle), présentées verticalement pour chacun des sujets des deux genres (moyennes sur 4 productions).

Par-delà ces différences interpersonnelles, la figure fait cependant apparaître des récurrences. On observe ainsi que souvent, le segment se réduit pratiquement à un point, les valeurs de début et fin de syllabe étant similaires sinon égales. Là où apparaît un segment de droite (c'est à dire en cas d'inégalité des valeurs initiale et finale), il est pratiquement toujours de faible ampleur en comparaison de la distance entre son point le plus bas et le cercle figurant la valeur de milieu de syllabe. Dans tous les cas, le cercle apparaît en-dessous du segment (ou du point), ce qui confirme, pour les individus considérés isolément, le constat global effectué plus haut (Figure 1) d'une infériorité du pitch de milieu de syllabe par rapport aux valeurs qui l'entourent.

Quelque intéressants que soient ces constats, ils ne permettent cependant pas d'accéder aux caractéristiques des réalisations isolément, puisque les données présentées dans la figure 2 sont des moyennes calculées sur 4 productions.

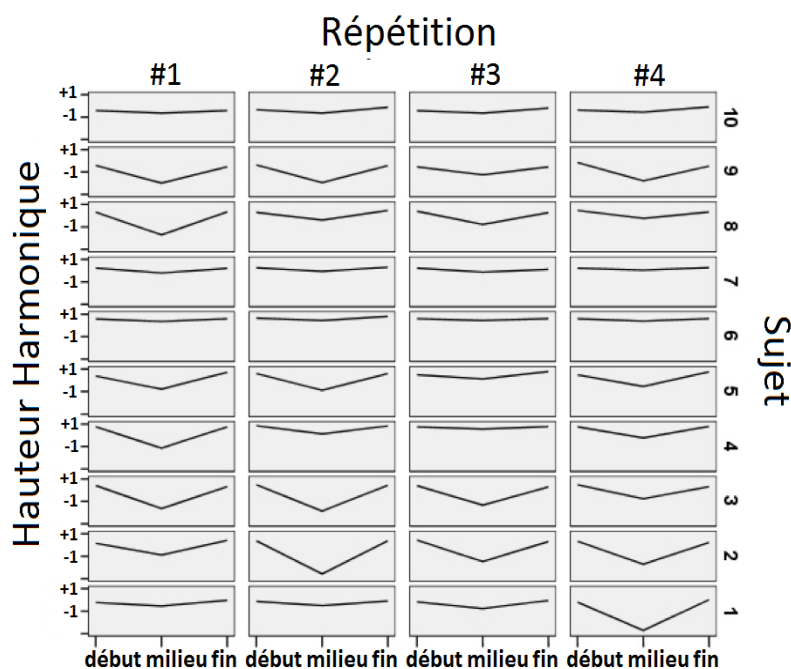


Figure 3 : évolution de la mélodie pour chaque production de syllabe, et chaque sujet (pitch en Hauteurs harmoniques).

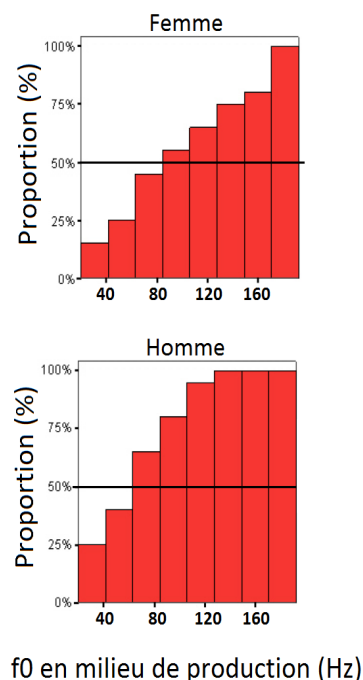


Figure 4 : distributions cumulées des valeurs centrales de f_0 (Hz): femmes et hommes.

Dans la figure 3, par contre, les profils d'évolution de la fréquence fondamentale dans la syllabe sont détaillés sujet par sujet et production par production. Afin de rendre plus comparables les profils tonaux (vu l'importante variabilité interindividuelle observée), les fréquences en Hertz (F) ont ici été converties en hauteurs harmoniques (Ha) selon la formule (Pierart B., Harmegnies B., 1993) :

$$Ha = (1/\text{LOG}_{10}(2)) * \text{LOG}_{10}(F/131)$$

Des profils de sujets très différents apparaissent ici. On observe, par exemple, que les sujets 2, 3, 8 et 9 opèrent systématiquement une diminution sensible du pitch en milieu de syllabe. A contrario, les sujets 6, 7 et 10 présentent des profils d'évolution pratiquement plats. Certains sujets, quant à eux, se montrent inconstants de production à production, affichant tantôt un profil plat, tantôt un profil en V.

3.3 Occurrences de (très) basses fréquences

Les données présentées plus haut indiquent la présence de fréquences très basses tant dans le groupe masculin que dans le groupe féminin. A ce propos, la figure 4 montre des histogrammes cumulés qui, construits au départ des valeurs individuelles de fréquences fondamentales de milieu de syllabe, rendent compte des valeurs de f_0 avec la plus fine des granularités possible. On peut y observer un nombre de très basses fréquences (même inférieures à 50 Hz) important et ce, tant chez les hommes que chez les femmes. La figure 5 illustre le phénomène en présentant le sonagramme d'une séquence /ma/ produite avec un ton T3 par un jeune locuteur. La fréquence fondamentale est de

l'ordre de 170 Hz tant en début qu'en fin de production, mais elle chute considérablement en milieu de production. Les pulses glottiques s'espacent et se raréfient graduellement, puis se resserrent tout aussi progressivement. Au centre, l'écart interpulses constaté est de 20 ms, ce qui correspond à une fréquence de 50Hz. Sur le plan perceptif, la zone centrale est systématiquement interprétée comme un épisode de voix craquée.

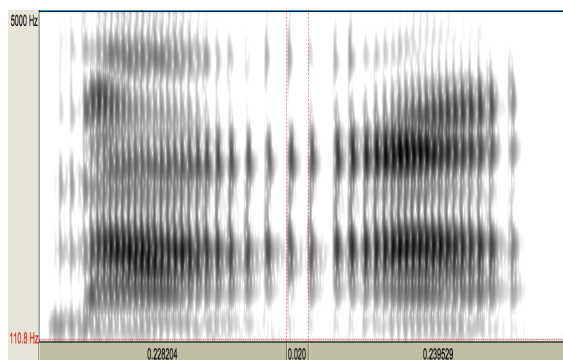


Figure 5 : sonagramme d'une réalisation de la syllabe /ma/ (T3), avec f_0 très basse (≈ 50 Hz).

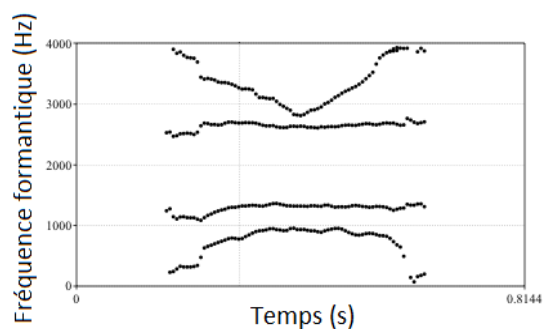


Figure 6 : graphe formantique de la réalisation de la syllabe /ma/ (T3, sujet 7, production #2), avec éclaircissement du timbre.

3.4 Autres corrélats acoustiques du T3

L'examen des valeurs des formants indique que certains sujets recourent (occasionnellement ou systématiquement) à un éclaircissement du timbre lors de la production du T3. Celui-ci est illustré dans le graphe formantique (Cf. Figure 6) par la production 2 du sujet 7 (celui de notre échantillon qui présente de la manière la plus nette et la plus systématique ce profil). Comme on peut le voir, c'est la dynamique du quatrième formant qui semble ici déterminante, sa fréquence passant de 2800 Hz en milieu de syllabe à 4100 Hz en fin de syllabe. Dans une étude récente (Zhang, 2017) sur les tons lexicaux du Chinois Mandarin en voix modale et en voix chuchotée, on observe que la dynamique du quatrième formant est parfois plus net, en liaison avec la nature du segment vocalique et le mode de phonation, que celui des trois premiers formants. Trois des sujets de notre échantillon présentent cette tendance (sujets 3, 7 et 10). Pour deux d'entre eux, les réalisations ne comportent en effet pratiquement pas de variation mélodique, ce qui suggère une stratégie compensatoire fondée sur la variation du timbre pour produire la dynamique nécessaire à l'implémentation du T3, qui repose ici plus sur l'évolution du contenu spectral que sur celle de la fréquence fondamentale.

Enfin, les mesures de durée syllabique effectuées sur l'ensemble du corpus (T1, T2, T3, T4) font apparaître des différences significatives de durée en fonction du ton, confirmées par l'analyse de variance ($F=4625,797$, $df=3$, $p<.001$). Pour chacun des 10 sujets, le ton 3 est systématiquement porté par une syllabe plus longue. Cette différence de durée est cependant plus marquée chez certains locuteurs que chez d'autres, ce qui pourrait suggérer l'existence de stratégies individuelles de production du T3 recourant à ce paramètre.

4 Discussion et Conclusion

Dans cette étude exploratoire, nous avons décrit acoustiquement la production du T3 au départ de productions isolées de mots et pseudo-mots monosyllabiques par 10 locuteurs natifs du chinois mandarin. En ce qui concerne l'évolution temporelle de la fréquence fondamentale, on retrouve globalement la forme de "V" à laquelle on pouvait s'attendre étant donné la transcription phonologique traditionnellement associée à ce ton (214). Notre étude aboutit cependant à nuancer ce constat sur deux points. Premièrement, les valeurs de fréquence fondamentale en début et fin de T3 divergent peu l'une de l'autre et ne justifient certainement pas un "saut" de deux registres (de 2 à 4). Deuxièmement, la forme en "V" résume adéquatement les données considérées dans leur ensemble mais ne caractérise pas chaque production réalisée par chaque locuteur.

Se pose dès lors la question des autres corrélats acoustiques potentiels de la réalisation du T3, outre l'évolution temporelle de la f_0 . Dans cette étude, nous en avons considéré trois. Tout d'abord, en accord avec la littérature récente (Kuang, 2017), nous avons régulièrement constaté la présence de fréquences extrêmement basses dans la phase médiane de la réalisation du ton, qui étaient associées sur le plan perceptif à un épisode de voix craquée. Il faut souligner, cependant, que nous n'avons observé ici aucune interruption de la périodicité, ni aucun bruit venant se surajouter au phénomène, ce qui caractérise généralement le type "creaky voice" (p.ex. Keating et al., 2015). La question du mécanisme phonatoire sous-jacent demeure donc à ce stade. Il ne semble pas possible en tout cas qu'une si faible fréquence fondamentale puisse être obtenue à l'aide des seuls plis vocaux. L'hypothèse qu'une structure plus massive est ici mobilisée, peut être en sus des plis (bandes ventriculaires ?) devrait être testée au moyen d'instrumentation articulatoire. On pourrait alors considérer la possibilité qu'un mode phonatoire particulier a été adopté par des membres d'une communauté linguistique donnée (ici, celle des mandarinophones) et que certains locuteurs ont acquis la capacité d'en faire usage avec plus ou moins de succès, mais loins d'être systématique, ce qui confirme l'étude de Moisik, Lin & Esling (2014).

En effet, un autre aspect des données analysées ici est la grande variabilité inter-individuelle observée dans l'utilisation d'indices acoustiques complémentaires susceptibles de supporter la réalisation phonétique du T3. Certains locuteurs présentent des épisodes de voix craquée, d'autres recourent à un éclaircissement du timbre, d'autres encore à une augmentation plus importante de la durée syllabique (ce qui n'est pas toujours le cas dans la production de la parole connectée, où la durée vocalique des tons est réduite quasiment tous à la même durée, le profil des tons modifié due à la transition aux frontières tonales. Cela crée donc la confusion entre certains paires de tons tels que la paire T2/T3), ce qui suggère une diversité de stratégies individuelles (compensatoires?) permettant de supporter phonétiquement le contraste phonologique entre le T3 et les autres tons du mandarin. Il importe de souligner ici que rien n'indique une quelconque exclusivité mutuelle des démarches stratégiques observées. D'une part, certains locuteurs paraissent tantôt recourir à une stratégie et tantôt pas (par exemple le sujet 1, qui ne recourt massivement à la stratégie mélodique que dans sa production 4). D'autre part, rien n'exclut que plusieurs stratégies puissent simultanément être utilisées par certains locuteurs dans certaines productions. Bien entendu, la question de l'efficacité de ces stratégies et / ou de leurs combinaisons, et celle d'une éventuelle hiérarchie entre les divers corrélats acoustiques ici considérés, devra être adressée au cours de futures études en perception de la parole.

Références

- BELOTEL-GRENIÉ, A., & GRENIÉ M. (1994). Phonation types analysis in standard Chinese. *Proceedings of Spoken Language Processing*, 343-346.
- BRUNELLE, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, 37, 79-96.
- CHAO, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique*, 45, 24-27.
- CHAO, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday*, édité par M. Halle, H. Lunt, H. McLean, & C. V. Schooneveld (Mouton Publishers, The Hague, The Netherlands), pp. 52-59.
- DAVISON, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics*, 78, 50-57.
- DUANMU, S. (2007). *The phonology of standard Chinese*. Oxford University Press.
- KEATING, P., GARELLEK, M., & KREIMAN, J. (2015). Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK, pp. 0821.1-0821.5.
- MOISIK, S.R., LIN, H., & ESLING, J.H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association*, 44(1), 21-58.
- KUANG, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, 142, 1693.
- PIERART, B., & HARMEGNIES, B. (1993). Dysphasie simple de l'enfant et langage de la mère. *L'année psychologique*, 93(2), 227-268.
- YANG, R.X. (2011). The phonation factor in the categorical perception of Mandarin tones. *Proceedings of ICPhS XVII*, Hong Kong, China, 2204-2207.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- YU, K.M., & LAM, H.W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, 136(3), 1320-1333.
- ZHANG, X.L. (2017). Les tons lexicaux du chinois mandarin en voix modale et en voix chuchotée. Thèse, 238.