

# Lire les tons sur les lèvres : perception(s) visuelle(s) des tons lexicaux en chinois mandarin

Grégory Roulet-Guiot<sup>1</sup> Corine Astésano<sup>2,3</sup>

(1) XJTU, Université Jiaotong, Xi'an, Chine

(2) LPL, UMR7309, Aix-en-Provence, France

(3) Octogone-Lordat, E.A. 4156, Toulouse, France

rouletguiotgregory@hotmail.fr, corine.astesano@univ-tlse2.fr

## RESUME

---

La présente étude a pour but de vérifier si les informations visuelles situées au niveau du cou peuvent contribuer à la perception visuelle des tons en mandarin. Cependant, ce que montre principalement cette étude est que les tons peuvent être lus sur les lèvres, et ce contre toute attente, même lorsque la syllabe est prononcée en arrière de la cavité buccale. En effet, il semblerait d'une part que la lecture labiale soit possible pour les tons du mandarin, et d'autre part qu'il existe différents profils de perception : certaines personnes semblent plus sensibles à la lecture labiale, alors que d'autres auraient *a priori* recours aux informations visuelles au niveau du cou. En contrepartie, ces personnes montreraient une aptitude moindre à la lecture labiale.

## ABSTRACT

---

### Read the tones on the lips : visual perception(s) of lexical tones in Mandarin Chinese

The aim of the present study is to verify whether the visual cues located on the neck, can contribute in Mandarin tones visual perception. However, in an unexpected way, this study shows that tones can be read on the lips, even when the syllable is pronounced in the back of the oral cavity. It seems indeed on the one hand that the labial reading is possible for Mandarin tones, on the other hand, that there could be various profiles of perception : some people seem to be more sensitive to the labial reading, other people could *a priori* use the neck's cues, and they would be less suited to the labial reading.

MOTS-CLES : chinois mandarin, tons, perception audiovisuelle, lecture labiale, multimodalité.

KEYWORDS : Mandarin Chinese, tones, audiovisual perception, labial reading, multimodality.

---

## 1 Introduction

La perception de la parole est reconnue depuis l'effet McGurk (McGurk & MacDonald, 1976) comme étant bimodale. En effet, si, dans une communication *de visu*, le canal auditif est suffisant pour décoder la parole, il est pour autant complété et influencé par le canal visuel. Cette découverte a donné lieu à de nombreuses études, et ce dans et entre différentes langues. Certaines études mettaient en avant que la magnitude de l'effet McGurk était très variable d'une langue à l'autre. Cependant, nous retenons l'étude de Massaro & al (1993), qui tend à prouver que la magnitude de l'effet McGurk est équivalente entre les langues, mais que les différences constatées sont le fait des contraintes phonotactiques intrinsèques à chaque langue. Depuis, les linguistes s'intéressant à la perception de la parole mettent non plus en avant la bimodalité de la parole, mais la multimodalité due à la multisensorialité dont Schwartz (2004) parle comme étant « au cœur de la communication parlée ».

La présente étude porte sur le chinois mandarin, plus précisément le *putonghua* (littéralement : « langue commune », il s'agit aujourd'hui de la langue véhiculaire dont s'est dotée la Chine, celle-ci est basée sur le mandarin de Pékin. Ci-après mandarin). Le mandarin est une langue tonale, à savoir que les tons lexicaux ont une valeur suprasegmentale distinctive. L'origine des tons est une variation de la F0 créée au niveau du système laryngé. La question est de savoir si certains stimuli visuels sont corrélés aux tons lexicaux. Des études récentes sur les langues à tons démontrent qu'ils existent des rapports entre les tons et certains indices visuels. En cantonnais et en thaïlandais il a été mis en évidence une relation entre les tons et les mouvements de la tête (Burnham et al, 2006). En mandarin, Chen & Massaro (2008) montrent qu'il est possible d'entraîner des sinophones natifs à reconnaître les différents tons en focalisant leur attention sur les informations visuelles corrélées aux quatre tons du mandarin. Parmi ces informations, ils ont relevé des mouvements différents au niveau du cou lors de la production des différents tons. Ces mouvements sont dus à l'action en partie visible de différents muscles (sterno-hyoïdien, sterno-thyroïdien et thyro-hyoïdien) permettant au larynx la production des tons, car lorsque le larynx s'abaisse cela implique un mouvement descendant de la F0 et inversement. Il a en effet été prouvé que ces mouvements du larynx sont corrélés avec les valeurs de la F0 (Honda & al, 1999). De plus, il semblerait qu'une coordination entre le système laryngé et le système articuloire fasse apparaître des indices visuels permettant de distinguer les tons entre eux (Burnham, 2000). Dans leur étude, Tong et Manwa (2011) montrent que les différents tons du cantonnais ont des réalisations articuloires significativement différentes pour les plosives bilabiales /p<sup>h</sup>a/ et /pa/, à savoir que la mâchoire est plus ou moins ouverte suivant les différents tons. Ces auteurs se réfèrent à une étude sur le mandarin montrant que la prononciation des différents tons du mandarin implique une différence au niveau du positionnement de la langue et de la mâchoire (Erickson et al, 2004). D'après les données de leur propre étude ainsi que celle d'Erickson et al., Tong et Manwa (2011) se positionnent alors contre la *source-filter theory* émise par Pickett (2001), que l'on peut résumer par une indépendance du système laryngé (source) par rapport au système supralaryngé (filtre). S'il était accepté jusqu'alors que la production des tons lexicaux était indépendante de l'articulation, les données susmentionnées montrent le contraire. Le système laryngé semble corrélé au système supralaryngé, et ce, au moins pour le cas des tons lexicaux.

Ces études tendent à montrer qu'il existe des corrélats visuels aux tons lexicaux. La F0 n'est donc pas la seule source d'information permettant le décodage des tons lexicaux. A ce propos, une étude de Liu & Samuel (2004) montre que même lorsque la F0 est neutralisée (la F0 est remplacée par du silence et est resynthétisée sous Praat), les locuteurs sont pourtant capables de discriminer les tons. D'une manière plus écologique, c'est aussi ce que montre l'étude de Chang & Yao (2007) sur la discrimination des tons en mandarin dans un contexte de parole chuchotée (donc sans l'indice de F0). Leur étude montre néanmoins que les locuteurs du mandarin peuvent se comprendre dans ce contexte. Les auteurs émettent le postulat que les locuteurs doivent s'appuyer sur la durée et l'intensité propres à chaque ton pour être capable de les reconnaître. Ce n'est *a priori* pas le cas. La multisensorialité de la parole aurait donc pour effet de démultiplier les indices lors du codage et du décodage.

Le but de la présente expérience est de vérifier si les informations visuelles situées au niveau du cou peuvent participer à la perception des tons dans une tâche de reconnaissance des tons sans le son. Nos hypothèses de départ sont : 1-que les informations visuelles situées au niveau du cou permettront une meilleure reconnaissance des tons ; 2-moins les participants

ont accès aux informations visuelles plus leur taux de reconnaissance devrait diminuer ; 3-du fait de la lecture labiale, les participants devraient montrer un taux de reconnaissance bien plus élevé des stimuli de la syllabe /p<sup>hi</sup>/, que des stimuli de la syllabe /g<sup>ŷ</sup>/ car la consonne bilabiale est plus visible que la consonne vélaire.

## 2 Matériel et méthode

### 2.1 Matériel linguistique

L'enregistrement des stimuli a été effectué dans la chambre sourde du laboratoire de phonétique de l'UQÀM (Université du Québec À Montréal). Le matériel utilisé pour les enregistrements est une caméra mini-DV Panasonic DVX100A pour la vidéo, et un micro unidirectionnel Audio-Technica ATM31a pour l'audio. Le taux d'échantillonnage est de 29,97 images/s pour la vidéo et de 22 kHz pour l'audio. Ces enregistrements ont été effectués sur un PC via le logiciel Adobe® Premiere® Pro. Deux participants ont été recrutés pour l'enregistrement des stimuli : un homme de 39 ans et une femme de 26 ans, sinophones natifs de Chine continentale, ne parlant aucun dialecte et aucune autre langue à tons.

Deux syllabes du mandarin ont été sélectionnées pour notre étude : /p<sup>hi</sup>/ et /g<sup>ŷ</sup>/. La syllabe /p<sup>hi</sup>/ est constituée d'une consonne bilabiale et d'une voyelle antérieure, fermée et non-arrondie, alors que la syllabe /g<sup>ŷ</sup>/ est constituée d'une consonne vélaire et d'une voyelle postérieure, mi-fermée et non-arrondie. Ces deux syllabes ont la particularité d'occuper les extrémités du système phonologique du mandarin sur l'axe antérieur/postérieur de la cavité buccale, nous permettant ainsi de contrôler la lecture labiale. Notre étude a pour objectif de mettre en évidence l'apport des stimuli visuels situés au niveau du cou dans le décodage de la parole dans un échange face à face. Afin de contrôler les différents indices visuels, nous avons choisi de réaliser 3 cadrages (les différents cadrages ont été réalisés après les enregistrements sur un ordinateur MacBook® à l'aide du filtre *blacken borders* du logiciel Avidemux2®, ce filtre ayant la particularité de pouvoir intégrer des bordures noires sur des vidéos, mais surtout de ne pas altérer ni la taille ni la qualité de la vidéo). Le cadrage témoin est un cadrage au niveau des épaules dans lequel on voit l'ensemble du visage. Nos deux cadrages expérimentaux sont un cadrage dans lequel on voit la bouche et le cou, et un cadrage dans lequel on ne voit que la bouche (cf. figure 1 ci-après).

Les deux syllabes à l'étude sont présentées dans 4 blocs différents, soit un total de 8 blocs. Les 4 blocs associés aux 3 cadrages, sont constitués d'un bloc témoin et de 3 blocs expérimentaux, tels que :

- Un bloc témoin : cadrage épaule avec le son appelé Avec Son (AS)
- Un bloc expérimental 1 : cadrage épaule sans le son appelé Sans Son (SS)
- Un bloc expérimental 2 : cadrage cou + bouche sans le son appelé Cadrage Cou (CC)
- Un bloc expérimental 3 : cadrage bouche sans le son appelé Cadrage Lèvre (CL)

Chaque bloc ne peut contenir que les stimuli d'une seule des deux syllabes à l'étude. Les stimuli de chaque syllabe correspondent à la prononciation de celle-ci avec les quatre tons du mandarin, et ce par 2 locuteurs différents. Ces stimuli sont répétés 3 fois, soit 4 tons X 2 locuteurs X 3 répétitions = 24 stimuli dans chaque bloc, présentés en ordre pseudo-aléatoire.

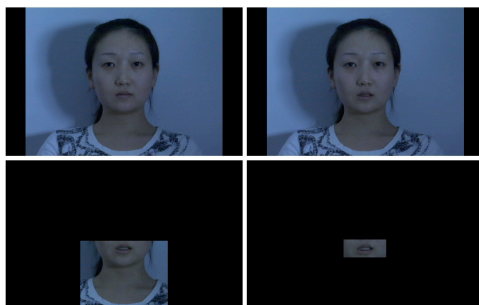


FIGURE 1 – Les quatre conditions expérimentales, soit trois types de cadrage : de gauche à droite et de haut en bas: AS, SS, CC, CL.

## 2.2 Participants

32 participants ont été recrutés pour cette expérience (âgés de 21 à 23 ans), sinophones natifs, en 3e et 4e année du département de français de l'université Jiaotong de Xi'an. Une partie de ces étudiants parle le dialecte de leur région d'origine. Aucun ne rapporte avoir de problème auditif. La majorité a une vision corrigée et portait des lunettes ou des lentilles de contact lors du test.

## 2.3 Procédure et déroulement de l'expérience

Nous reprenons pour notre expérience une partie de la méthodologie de Chen & Massaro (2008), à laquelle nous ajoutons un contexte expérimental : un cadrage dans lequel on voit la bouche et le cou. Notre but est de montrer que les informations visuelles situées au niveau du cou participent à la discrimination tonale dans un contexte uniquement visuel (sans le son). Notons que les participants ne sont pas entraînés à ce test, car il s'agit ici de connaître l'influence des informations visuelles dans un contexte proche du contexte écologique. Pour information, s'agissant d'une étude préliminaire, le temps de réaction n'est pas pris en compte dans cette expérience : seul le pourcentage de réponses correctes est calculé.

Les blocs sont présentés avec le logiciel PowerPoint® : chaque stimulus est précédé d'une diapositive portant le numéro du stimulus, permettant d'indiquer au participant à quelle ligne les stimuli correspondent sur la feuille-réponse. La feuille-réponse est constituée de 24 lignes numérotées sur lesquelles sont imprimés des caractères chinois correspondant aux quatre tons du mandarin. Les graphies retenues sont des graphies canoniques du mandarin.

Le déroulement du test est le suivant: 1. Les participants ne peuvent voir chaque stimulus qu'une seule fois ; 2. cependant, comme ils contrôlent eux-mêmes le passage d'un stimulus à l'autre (en appuyant sur la touche « → » du clavier), ils peuvent prendre le temps dont ils ont besoin pour entourer le caractère qui correspond à ce qu'ils ont vu ou entendu. Le test se déroule en 4 parties : 1. un pré-test avec la syllabe /ma/ servant d'entraînement à la tâche avec uniquement les stimuli de la locutrice, présentant une articulation plus nette ; 2. la présentation des blocs d'une des deux syllabes. Notons qu'un ordre de présentation des blocs est attribué à chaque participant dès le début en gardant toujours le bloc témoin avec

le son (AS) en premier (par ex. : AS, CL, SS, CC) : un participant aura donc un ordre de présentation des blocs identique pour l'ensemble du test (pré-test ; 1<sup>ère</sup> syllabe ; 2<sup>ème</sup> syllabe). De plus, l'ordre de présentation des syllabes est aussi aléatoire ; 3. Une période de pause d'environ 10 minutes est prévue au milieu du test. De plus, les participants disposent d'une courte pause entre chaque bloc ; 4. La présentation des blocs de l'autre syllabe.

### 3 Résultats

Nous choisissons de ne présenter dans cette étude préliminaire que les résultats concernant la locutrice, car son articulation est plus claire que celle du locuteur, et les taux de réponses correctes associés aux différentes conditions expérimentales sont significativement plus élevés que pour le locuteur. Les participants ayant le choix entre 4 réponses différentes (4 tons), le pourcentage de réponse attribué au hasard pour ce test est de 25% : les résultats de se situant au-dessus de 25% peuvent donc donner lieu à des interprétations. Les résultats présentés ci-dessous concernent les scores significativement supérieurs à 25%. Les données sont analysées à l'aide du *t* de *student* (avec un seuil de significativité à  $p < .05$ ).

#### 3.1 Différence inter-syllabe

Les résultats dans la condition Avec Son (AS, condition témoin), sont très élevés. Pour la syllabe /gʏ/ la moyenne des participants est de 98% de reconnaissance des tons, les tons de la syllabe /p<sup>hi</sup>/ sont quant à eux reconnus à 96% dans la condition AS. Ces résultats témoignent de la qualité de nos stimuli et de la capacité des participants à les percevoir.

Nos résultats montrent une différence significative entre les deux syllabes dans chacune des conditions expérimentales (SS :  $p < .002$  ; CC :  $p < .005$  ; CL :  $p < .006$ ). Globalement, la moyenne du taux de reconnaissance à travers les 4 conditions (AS, SS, CC, et CL), est de 50,4% pour la syllabe /gʏ/, et de 55,1% pour la syllabe /p<sup>hi</sup>/ . L'ensemble de nos données montre donc un taux de reconnaissance plus important pour la syllabe /p<sup>hi</sup>/ que pour la syllabe /gʏ/, cette différence étant significative dans les 3 conditions expérimentales. La moyenne du taux de reconnaissance à travers les trois conditions expérimentales (SS,CC et CL), pour la syllabe /gʏ/ est de 34,3%, alors qu'elle est de 41,7% pour la syllabe /p<sup>hi</sup>/ . Tout porte à croire que l'antériorité de la syllabe est en cause dans cette différence.

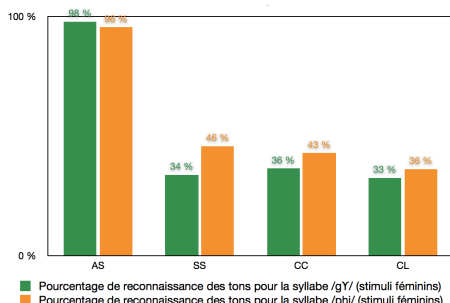


FIGURE 2 – Représentation graphique du taux de reconnaissance des tons pour les stimuli féminins dans les quatre contextes. En vert la syllabe /gʏ/, en orange la syllabe /p<sup>hi</sup>/ .

Pour la syllabe /gʏ/, nos résultats n'indiquent pas de différence significative ( $p > .05$ ) entre les différentes conditions expérimentales (SS, CC et CL). Notons toutefois une tendance (non significative) de meilleure reconnaissance pour la condition CC (cf. figure 2, ci-dessus). Les taux de reconnaissance de la syllabe /p<sup>hi</sup>/ sont en revanche significativement plus élevés en CC qu'en CL ( $p < .02$ ) et en SS qu'en CL ( $p < .004$ ). Alors qu'on aurait pu s'attendre à une influence très supérieure des informations visuelles du cou dans le taux de reconnaissance des tons (action des muscles lors de l'abaissement/élévation du larynx en fonction du ton), nos données indiquent que la lecture labiale a également une incidence sur le décodage des mouvements tonals, notamment pour les syllabes antérieures.

### **3.2 Résultats interparticipants : une variabilité importante, des aptitudes exceptionnelles.**

Nous nous intéresserons maintenant à la capacité que présentent certains participants à très bien reconnaître les tons sur la base de la lecture labiale, y compris dans le contexte peu visible de la syllabe /gʏ/. Nous affinons donc nos résultats en regroupant les participants dont le taux de reconnaissance était élevé en CL d'un côté, et ceux dont le taux de reconnaissance était faible en CL de l'autre. Nous retenons un seuil de 40% de reconnaissance, assez éloigné du seuil du hasard.

#### **3.2.1 Syllabe /p<sup>hi</sup>/**

En regroupant les participants suivant leurs résultats (>40%) dans la condition Cadrage Lèvres (CL, information visuelle minimale), 12 participants (37,5% des effectifs) présentaient une moyenne élevée allant de 50,7% dans la condition CL, à 51,4% dans la condition CC (45,8% en SS). Il semblerait que les participants qui ont une capacité à la lecture labiale aient un taux de reconnaissance élevé dans tous les contextes expérimentaux.

En regroupant les participants suivant leurs résultats (>40%) dans la condition CC, 20 participants sur les 32 participants (soit 62,5% des effectifs) obtiennent 52,5% de reconnaissance dans la condition CC, et 50% dans la condition SS. Cependant, ces personnes montrent un taux de reconnaissance en deçà de 40% dans la condition CL (38,8%). Les personnes ayant un taux de reconnaissance important dans les conditions SS et CC présentent donc un taux de reconnaissance assez faible dans la condition CL. Ont-ils alors recours à d'autres informations visuelles que celles situées au niveau des lèvres ? S'agit-il des informations apportées par les mouvements du larynx au niveau du cou ?

#### **3.2.2 Syllabe /gʏ/**

Treize participants (soit 40% des effectifs) ont un taux de reconnaissance élevé dans la condition CL (46,2%). Leurs résultats dans les conditions SS (30,8%) et CC (34%) sont légèrement inférieurs à ceux de l'ensemble des participants. Il est étonnant de noter que pour cette syllabe vélaire postérieure, les participants ont de meilleurs taux de reconnaissance en lecture labiale qu'en lecture des indices du cou. On peut émettre l'hypothèse que ce type de participants qui utilisent davantage l'information labiale ne soit pas en mesure d'utiliser d'autres indices, pourtant plus visibles pour la syllabe vélaire postérieure.

En regroupant les participants qui ont un taux de reconnaissance élevé dans la condition CC,

on constitue un groupe de 17 participants (53% des effectifs) pour un taux de reconnaissance moyen de 47,1% dans la condition CC. Les participants de ce groupe montrent la même disposition que ceux du groupe de la syllabe /p<sup>hi</sup>/, à savoir qu'il obtiennent de meilleurs résultats dans la condition CC que dans la condition CL (32,4%), mais aussi que dans la condition SS (36,8%).

## 4 Discussion

La présente étude se veut préliminaire, vu la complexité de l'analyse des données. Le nombre de locuteurs doit idéalement être augmenté afin de mieux expliquer, en diminuant le « bruit » statistique, la variabilité observée entre les deux locuteurs enregistrés pour la présente étude. Afin de rendre compte également de la tendance que nos résultats mettent en évidence selon laquelle il existerait deux stratégies de décodage visuel des tons (prédominance de la lecture labiale vs. prédominance de l'utilisation d'indices venant de la position du larynx), il serait nécessaire d'augmenter le nombre de nos informateurs.

### 4.1 Lire les tons sur les lèvres

Les résultats de cette étude mettent en avant le fait que la lecture labiale est possible en chinois mandarin. C'est ce que tend à prouver une étude récente (Tong et Manwa, 2011) pour le cantonnais dans le cadre des bilabiales /p<sup>ba</sup>/ et /pa/. En 2004, une étude (Erickson et al) avait prouvé que les tons du mandarin donnaient lieu à des positions différentes de la langue et de l'ouverture de la bouche, corrélées avec les tons. Cette étude portait aussi sur des syllabes bilabiales (/ba/, /ma/, /pa/). Conformément à notre hypothèse, on a pu montrer une différence significative dans la reconnaissance des tons entre les deux syllabes à l'étude : /p<sup>hi</sup>/ bilabiale antérieure mieux identifiée que /g<sup>ŷ</sup>/ vélaire postérieure, et ce dans l'ensemble des contextes expérimentaux. Pour autant, nos résultats semblent montrer que pour le mandarin, la lecture labiale des tons est aussi possible pour des syllabes postérieures telles que /g<sup>ŷ</sup>/ . Ce qui soutient et renforce l'hypothèse de la lecture labiale des tons lexicaux.

### 4.2 Différentes stratégies de décodage visuel des informations tonales ?

De manière générale, nos données montrent des taux de reconnaissance individuels des tons sans le son de près de 60% dans chacun des contextes expérimentaux des deux syllabes. De plus, si certains participants montrent un taux de reconnaissance plus élevé en condition Cadrage Lèvres (CL, informations visuelles minimales) que dans les conditions Cadrage épaule Sans Son (SS, informations visuelles maximales) et Cadrage Cou (CC, informations visuelles intermédiaires), où les cadrages sont pourtant plus larges, d'autres montrent en revanche une reconnaissance relativement faible des tons dans les conditions expérimentales SS et CL, alors que leur taux de reconnaissance dans la condition Cadrage Cou (CC) est élevé, et ce, quelle que soit la syllabe présentée (/p<sup>hi</sup>/ antérieure et bilabiale vs /g<sup>ŷ</sup>/ postérieure et vélaire). Nos résultats révèlent donc deux stratégies potentielles de décodage visuel des informations tonales. On pourrait émettre l'hypothèse que certaines personnes utiliseraient des informations plus « globales » que d'autres, qui utiliseraient davantage la lecture labiale. Dans un choix forcé où le cadrage est focalisé sur les lèvres, il serait intéressant de pouvoir mesurer le rôle de la focalisation sur les informations visuelles. Ainsi, des informations données par le suivi du regard (*eyetracking*) seraient fondamentales

pour déterminer quels aspects sont davantage utilisés dans le décodage visuel de la parole, et nous permettraient de comprendre quelles stratégies sont mises en œuvre par les deux groupes de participants que nos résultats mettent en évidence, à savoir ceux qui ont de meilleurs taux de reconnaissance en cadrage large vs en cadrage restreint, et *vice versa*. Enfin, les résultats les plus inattendus de notre étude montrent une aptitude à la lecture labiale des tons lexicaux pour des syllabes prononcées à l'arrière de la cavité buccale. Aussi, il nous semble que cet aspect devrait être étudié via l'articulographie électromagnétique (EMA ; voir Erickson *et al*, 2004 ; Tong et Manwa, 2011 pour les bilabiales).

## Remerciements

Nous remercions Lucie Ménard (directrice du laboratoire de phonétique de l'UQÀM), d'avoir mis à notre disposition le laboratoire de phonétique de l'UQÀM pour effectuer l'enregistrement de nos stimuli, ainsi que toute l'équipe du laboratoire, et plus particulièrement Marilène C. Rousseau, pour nous avoir assistés lors des enregistrements. Aussi, nous tenons à remercier Niu Fan et Yong Gang (les locuteurs qui ont prêté leur voix permettant la réalisation de cette étude), ainsi que l'ensemble des participants.

## Références

- BURNHAM, D., CIOCCA, V., LAUW, C., LAU, S., et STOKES, S. (2000). Perception of visual information for Cantonese tones. *Actes de SST (Speech Science and Technology)*, Canberra.
- CHANG, C., et YAO, Y. (2007). Tone production in whispered Mandarin. *UC Berkeley Phonology Lab Annual Report*, pages 326-329.
- CHEN, T. H., MASSARO, D. W. (2008). Seeing pitch : Visual information for lexical tones of Mandarin-Chinese. *Journal of Acoustical Society of America*, pages 2356-2366.
- ERICKSON, D., IWATA, R., ENDO, M., et FUJINO, A. (2004). Effect of tone height on jaw and tongue articulation in Mandarin Chinese. *Proc. Tonal aspects of languages*, Beijing.
- HONDA, K., HIRAI, H., MASAKI, S., et SHIMADA, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech*, pages 401-411.
- LIU, S. F.; SAMUEL, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, pages 109-138.
- MASSARO, D. W., COHEN, M. M., GESI, A., HEREDIA, R., et TSUZAKI, M. (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics*, pages 445-478.
- MCGURK, H., et MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, pages 746-748.
- PICKETT, J.M. (2001). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. MA: Allyn and Bacon.
- SCWARTZ, J.-L. (2004). La parole multisensorielle: Plaidoyer, problèmes, perspective. *Actes des XXVes Journées d'Etude sur la Parole*.
- TONG, E. et MANWA, L. N. (2011). Interaction between lexical tone and labial movement in cantonese bilabial plosive production. *Actes de ICPhs XVII*, pages 2014-2017.