

Production conjointe de gestes brachio-manuels et de focalisation prosodique : coordination temporelle et effets de la production d'un geste sur les corrélats acoustiques/articulatoires de la focalisation

Benjamin Roustan et Marion Dohen

Département Parole et Cognition – GIPSA-lab – UMR5216 CNRS, Université de Grenoble
961, rue de la Houille Blanche 38402 Saint-Martin-d'Hères, France

ABSTRACT

Speech, and prosody in particular, is tightly linked to manual gestures. This study investigates the coordination of prosodic contrastive focus and different manual gestures (pointing, beat and control gestures). We used motion capture on ten speakers to explore this issue. The results show that prosodic focus "attracts" the manual gesture whichever its type. The temporal alignment between speech and the manual gesture is the strictest for pointing. It is realized between the pointing apex and an articulatory vocalic target whatever the position of focus within the utterance. Moreover, the results show that the production of a manual gesture, whichever its type, does not affect the acoustic and articulatory correlates of prosodic focus.

1. Introduction

Gestes manuels et parole sont liés dans la production de la parole en interaction (*e.g.* 1, 2). Plusieurs études [3–9] ont notamment mis en évidence les liens entre indices prosodiques et gestes manuels au niveau de la coordination gestes manuels/parole.

La deixis est la capacité d'attirer l'attention, de désigner. Dans l'espace, elle peut-être réalisée par le pointage manuel [10]. Dans la parole, la focalisation peut jouer ce rôle. En particulier, la focalisation contrastive prosodique est utilisée pour mettre en avant (*i.e.* désigner) un mot ou un groupe de mots au niveau informationnel. La deixis peut ainsi être réalisée de façon multimodale. Plusieurs études se sont intéressées à la coordination temporelle des réalisations unimodales de la deixis. De Ruyter [11] a montré que la variation de la position de l'accent lexical dans un mot simple n'a aucun effet sur la coordination temporelle parole/pointage. Par contre, il a montré que la variation de la position de la focalisation prosodique au sein d'un syntagme nominal (adjectif+nom) avait un effet sur l'instant d'initialisation du geste de pointer. Rochet-Capellan *et al.* [12] ont exploré la coordination parole/pointage dans une tâche combinée (pointer+parole) de dénomination d'une cible (non-mot de 2 syllabes) et en faisant varier la position de l'accentuation lexicale. Ils ont trouvé que la position de l'accent lexical avait une influence sur la coordination temporelle parole/pointage en ce sens que le geste était déplacé pour que la syllabe accentuée soit toujours contenue dans la partie du geste qui désigne (index maintenu pointé vers la cible). Il semblerait

donc que les différentes modalités de la deixis puissent être liées dans leur réalisation. Ce lien doit pourtant encore être précisé. Soulignons de plus que les études citées ci-dessus se sont principalement intéressées à des productions vocales simples (non-mots, mots ou syntagmes nominaux isolés).

Plusieurs études (*e.g.* 1, 5) ont de plus mis en évidence le lien potentiel entre prosodie et gestes de battements (*beat gestures* ou *batons* : oscillation verticale de la main de haut en bas).

Outre la coordination temporelle gestes manuels/parole, se pose également la question de l'influence de la production de gestes manuels sur la parole. Krahmer et Swerts [13] ont montré que la production d'un "battement visuel" (geste manuel, mouvement de sourcil ou hochement de tête) avait un effet significatif sur la durée des productions vocales concomitantes et sur le formant F_2 . Cet effet était par ailleurs proche des corrélats acoustiques d'une accentuation emphatique. Ce résultat suggère que la production d'un geste manuel aurait un effet sur la production de la parole concomitante. L'objectif de cette étude est d'analyser la coordination entre la production de la focalisation contrastive prosodique en français et plusieurs types de gestes manuels de natures différentes en utilisant des phrases complètes et des mesures de capture de mouvement. Les questions posées sont les suivantes : 1. La focalisation prosodique et les gestes manuels sont-ils coordonnés temporellement et si oui, comment ? 2. Cette coordination éventuelle est-elle influencée par le type de geste et notamment son lien fonctionnel avec la parole ? 3. La production d'un geste manuel a-t-elle un effet sur les corrélats acoustiques/articulatoires de la focalisation prosodique ?

2. Méthodologie

2.1. Protocole expérimental

Corpus Le corpus était constitué de 4 phrases de structure syntaxique sujet (S) – verbe (V) – objet (O) en français (ex : Mumu tient le bébé). La structure syllabique était la suivante : S = 2 syl (nom propre) ; V = 1 syl (verbe d'action au présent) ; O = 1+2 syl (article+nom commun). Les mots cibles (S et O) commençaient tous par une consonne bilabiale.

Conditions expérimentales Nous nous sommes intéressés à deux conditions de focalisation : focali-

sation sur le sujet (FS; exemple : MUMU_F tient le bébé) et sur l'objet (FO). Deux conditions de parole ont été étudiées : parole seule et parole+geste manuel. Trois types de gestes manuels ont été analysés : pointage (geste manuel déictique communicatif), battement (geste manuel non déictique communicatif), et contrôle (appui sur un bouton; geste manuel non déictique non communicatif).

Tâches Une tâche de correction (voir 14) a été utilisée pour induire la production de focalisation prosodique dans un contexte naturel de dialogue. Les participants entendaient un prompt audio dans lequel deux locuteurs discutaient et devaient corriger la phrase du deuxième locuteur en fonction de ce qu'avait dit le premier. Il était simplement demandé aux participants d'effectuer la tâche de correction dans la condition demandée. Deux images relatives au dialogue entendu étaient affichées sur un écran en face du participant qui, en condition de pointage, pointait vers l'image adéquate en même temps qu'il effectuait la correction. En condition parole+geste manuel, la seule indication donnée aux participants était de produire le geste manuel en même temps qu'ils parlaient.

Protocole L'expérience était divisée en quatre blocs (un par condition : parole seule, parole+geste manuel avec trois types de gestes). Avant chaque bloc, les participants s'entraînaient brièvement à la tâche avec des phrases différentes de celles utilisées pendant les phases de test. L'ordre des blocs et l'ordre des stimuli à l'intérieur d'un bloc étaient aléatoires et différents pour chaque participant. Chaque bloc comportait 16 essais (4 phrases, 2 types de focalisation, 2 répétitions).

2.2. Participants

Dix adultes droitiers et de langue maternelle française ont participé à l'expérience (8 femmes et 2 hommes; âge moyen : 30, 2)

2.3. Dispositif Experimental

Les participants étaient assis sur une chaise devant un écran. Une position de repos était marquée sur une table située à leur droite. Il leur était demandé de placer leur index sur ce repère en phase de repos : de partir de cette position pour faire un geste puis d'y revenir une fois le geste terminé. Les mouvements de leurs lèvres et de leur main droite étaient enregistrés grâce à un système 3D de suivi du mouvement (Optotrak). Quatre diodes étaient placées sur leurs lèvres (une à chaque commissure, une au milieu de la lèvre supérieure et une au milieu de la lèvre inférieure). Trois autres diodes étaient placées sur leur main droite (2 sur l'index : une sur l'ongle et une sur la première phalange et une sur le dos de la main). Les productions vocales des locuteurs étaient enregistrées grâce à un microphone.

2.4. Mesures

Toutes les productions acoustiques ont été validées pour vérifier que les participants avaient bien produit la focalisation sur l'élément souhaité. Cette validation acoustique a été réalisée en vérifiant que les corré-

lats acoustiques de la focalisation prosodique (*e.g.* 14) avaient bien été produits. Les erreurs de production ont été exclues de l'analyse. Les frontières acoustiques des syllabes ont été étiquetées à l'aide du logiciel Praat [15]. Les maxima de fréquence fondamentale (F_0) et d'intensité (Int) correspondant à l'élément focalisé (S ou O) ont été également étiquetés. La durée (Dur) de l'élément focalisé a été calculée. L'ouverture des lèvres et la protrusion de la lèvre supérieure ont été extraites des données de suivi du mouvement (Optotrak). Les cibles vocaliques (CV_1 , CV_2) correspondant à chacune des voyelles des syllabes de l'élément focalisé (2 voyelles) ont ainsi pu être annotées (maxima d'ouverture des lèvres ou de protrusion). Concernant le mouvement du doigt, l'apex (P_A) et le début du geste de retour (l'index repart de sa position d'apex pour revenir vers la position de repos; P_R) ont été annotés. Pour le geste de pointage, l'apex correspond à la position la plus étendue de l'index vers la cible. Pour le geste de battement, l'apex a été identifié comme étant le point vertical le plus bas du mouvement. Pour le geste de contrôle, l'apex a été identifié comme étant le moment où l'index appuie sur le bouton. Les instants de réalisation de chacun de ces événements (maxima de F_0 et d'intensité, P_A et P_R) ont été normalisés sur la durée totale de l'énoncé afin d'éliminer la variabilité due aux différences segmentales des énoncés ou celle liée aux temps de réponse variables.

Toutes les variables dépendantes ont été testées en utilisant des ANOVAs à mesures répétées avec deux facteurs intra-participants : type de focalisation (2 niveaux : FS et OF) et condition gestuelle (pour les variables gestuelles *i.e.* P_A et P_R : 3 niveaux : pointage, battement et contrôle; pour les variables acoustiques et articulatoires *i.e.* F_0 , Int, Dur, CV_1 et CV_2 : 4 niveaux : parole seule + 3 types de gestes).

3. Résultats

3.1. Timings : Coordination temporelle parole/geste

Résultats généraux La Table 1 donne les résultats des analyses statistiques sur les instants d'occurrence de P_A , P_R , F_0 , Int, CV_1 , CV_2 (resp. t_{P_A} , t_{P_R} , t_{F_0} , t_{Int} , t_{CV_1} , t_{CV_2}).

Table 1: Résultats des ANOVAs sur les instants d'occurrence des événements acoustiques, articulatoires et gestuels

	Condition Focalisation	Condition Geste
t_{P_A}	$F(1, 9) = 114.4$, $p < .001$	$F(2, 18) = 24.3$, $p < .001$
t_{P_R}	$F(1, 9) = 99.5$, $p < .001$	$F(2, 18) = 0.6$, $p = .55$
t_{F_0}	$F(1, 18) = 1571.6$, $p < .001$	$F(3, 27) = 1.6$, $p = .21$
t_{Int}	$F(1, 18) = 2478.6$, $p < .001$	$F(3, 27) = 5.6$, $p = .01$
t_{V_T1}	$F(1, 18) = 3746.1$, $p < .001$	$F(3, 27) = 1.1$, $p = .36$
t_{V_T2}	$F(1, 18) = 2655.7$, $p < .001$	$F(3, 27) = 2.2$, $p = .11$

Étude du geste manuel — Le type de focalisation a un effet significatif sur les instants de réalisation de P_A . Le geste a tendance à être réalisé plus tard au sein de l'énoncé quand c'est l'objet qui est focalisé (FO) : on peut dire que la focalisation attire le geste. La condition gestuelle a également un effet significatif sur l'instant de réalisation de P_A mais pas sur celui

de P_R . Ceci suggère que les différents types de gestes ne sont pas réalisés de la même façon (en tout cas en ce qui concerne leur apex).

Étude de la parole — Le *type de focalisation* a un effet significatif sur toutes les variables acoustiques et articulatoires. Ceci correspond au fait que les corrélats acoustiques et articulatoires ont été mesurés sur S pour FS et sur O pour FO. Ils arrivent donc forcément plus tard en condition FO. La *condition gestuelle* n'a d'effet significatif sur aucune des variables considérées. La production d'un geste n'a donc aucun effet sur l'organisation temporelle interne de l'énoncé.

Alignements temporels On peut dire que deux points sont alignés dans le temps si la différence entre leurs instants de réalisation est proche de 0. Dans le but d'étudier l'alignement potentiel des gestes manuels avec la parole, nous avons calculé, pour chaque énoncé, les différences entre les instants de réalisation des événements gestuels (t_{P_A} et t_{P_R}) et les instants de réalisation des événements acoustiques (t_{F_0} et t_{Int}) et articulatoires (t_{CV_1} et t_{CV_2}). Nous avons ensuite calculé la moyenne de ces différences sur tous les énoncés pour chaque participant. La Figure 1 donne les résultats sur le calcul des moyennes et écart-types sur tous les participants (si une boîte est proche de zéro, les deux événements sont proches dans le temps).

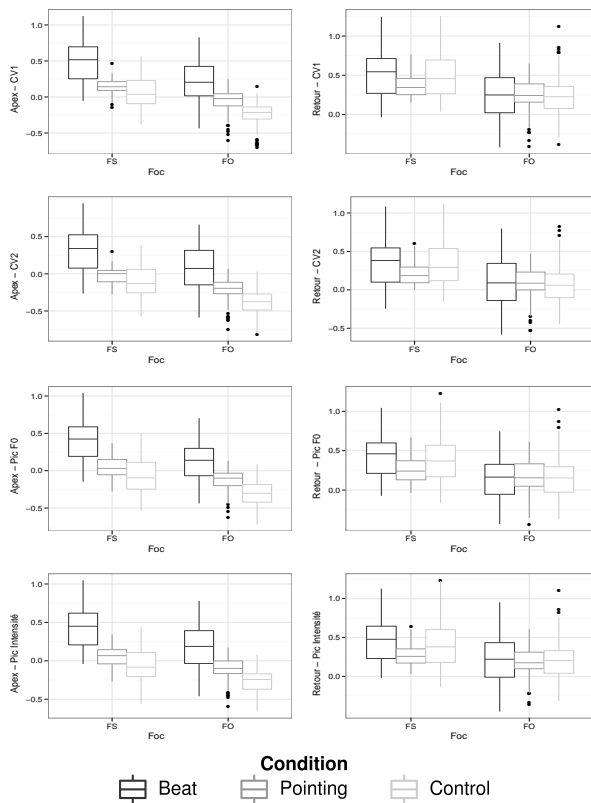


Figure 1: Alignements temporels entre les événements gestuels (P_A et P_R) et les événements acoustiques (F_0 , Int) et articulatoires (CV_1 , CV_2) pour toutes les conditions de focalisation (FS et FO) et tous les types de geste (pointage, battement, contrôle). Données temporelles normalisées (c.f. section 2.4)

La Table 2 donne les résultats des ANOVAs sur les

alignements temporels. La *condition de focalisation* a un effet significatif sur toutes les différences temporelles. Quel que soit le type de geste, la coordination temporelle entre la parole et les gestes manuels au sein de l'élément focalisé est différente pour FS et FO. La *condition gestuelle* a également un effet significatif sur toutes les différences temporelles pour P_A mais pas pour P_R . Les différents gestes manuels sont donc coordonnés à la focalisation prosodique de façons différentes : P_R se trouve à peu près au même endroit relativement aux corrélats acoustiques et articulatoires mais pas P_A .

Table 2: Résultats des ANOVAs sur les différences temporelles

	Condition Focalisation	Condition Geste
$t_{P_A} - t_{F_0}$	$F(1, 9) = 47.8, p < .001$	$F(2, 18) = 25.1, p < .001$
$t_{P_A} - t_{Int}$	$F(1, 9) = 55.3, p < .001$	$F(2, 18) = 25.0, p < .001$
$t_{P_A} - t_{VT_1}$	$F(1, 9) = 57.5, p < .001$	$F(2, 18) = 24.4, p < .001$
$t_{P_A} - t_{VT_2}$	$F(1, 9) = 55.6, p < .001$	$F(2, 18) = 24.4, p < .001$
$t_{P_R} - t_{F_0}$	$F(1, 9) = 32.8, p < .001$	$F(2, 18) = 0.69, p = .51$
$t_{P_R} - t_{Int}$	$F(1, 9) = 32.9, p < .001$	$F(2, 18) = 0.95, p = .40$
$t_{P_R} - t_{VT_1}$	$F(1, 9) = 38.6, p < .001$	$F(2, 18) = 0.61, p = .55$
$t_{P_R} - t_{VT_2}$	$F(1, 9) = 36.3, p < .001$	$F(2, 18) = 0.64, p = .53$

Des tests t (Welch) ont ensuite été menés pour comparer les instants normalisés d'occurrence des événements gestuels (P_A et P_R) aux événements acoustiques (F_0 et Int) et articulatoires (CV_1 et CV_2). Ces comparaisons ont été effectuées pour chaque type de geste séparément (puisque'il y a un effet de la condition gestuelle) et pour chaque type de focalisation séparément (puisque'il y a un effet du type de focalisation).

Geste de pointage — Les tests ont montré que t_{P_A} n'était pas significativement différent de t_{F_0} ($t(9)=1$; $p=0,3$) et de t_{Int} ($t(9)=1,5$; $p=0,2$) pour FS mais pas pour FO (t_{F_0} : $t(9)=-3,11$; $p=0,008$; t_{Int} : $t(9)=-2,7$; $p=0,02$). Pour FS, t_{P_A} n'est pas significativement différent de t_{CV_2} ($t(9)=-0,8$; $p=0,4$) et pour FO, t_{P_A} n'est pas significativement différent de t_{CV_1} ($t(9)=-1,5$; $p=0,2$). Il apparaît donc que pour le pointage, il y a alignement entre l'apex du geste et une cible articulatoire de l'élément focalisé.

Geste de battement — Pour FS, il n'y a aucun alignement entre P_A et un des corrélats acoustiques ou articulatoires de la focalisation. Pour FO, P_A semble être aligné avec les maxima de F_0 et d'intensité (F_0 : $t(9)=1,4$; $p=0,2$; Int : $t(9)=-0,8$; $p=0,4$) ainsi qu'avec CV_2 ($t(9)=0,7$; $p=0,5$).

Geste de contrôle — Pour FS, P_A semble être aligné avec les maxima de F_0 et d'intensité (F_0 : $t(9)=-1,2$; $p=0,3$; Int : $t(9)=-0,8$; $p=0,4$). Ceci n'est cependant pas le cas pour FO. Pour FS, P_A est aligné avec CV_1 ($t(9)=0,8$; $p=0,4$). Pour FO, P_R est aligné avec CV_2 ($t(9)=1,2$; $p=0,3$). De façon très intéressante, la Figure 1 montre aussi que les alignements sont plus précis pour le geste de pointage (voir les écart-types). Ceci est d'autant plus vrai si on regarde les données concernant l'apex.

3.2. Réalisations acoustiques et articulatoires de la focalisation : effets de la production d'un geste manuel et du type de geste

Nous avons analysé les amplitudes des corrélats acoustiques (durée de l'élément focalisé et maxima de F_0

et d'intensité) et articulatoires (CV_1 et CV_2). Le *type de focalisation* a un effet significatif sur toutes les variables. Des analyses post-hoc révèlent que les amplitudes de tous les corrélats acoustiques et articulatoires de la focalisation sont plus faibles en FO qu'en FS ce qui n'est pas surprenant [14]. De façon plus intéressante, la *condition gestuelle* n'a d'effet significatif sur aucune des variables : la production d'un geste n'affecte ni les corrélats acoustiques ni les corrélats articulatoires de la focalisation prosodique.

Table 3: Résultats des ANOVAs sur les corrélats acoustiques et articulatoires de la focalisation

	Condition Focalisation	Condition Geste
Dur	$F(1, 9) = 13.5, p = .05$	$F(3, 9) = .4, p = .7$
F_0	$F(1, 9) = 17, p < .01$	$F(3, 9) = 2.7, p = .1$
Int	$F(1, 9) = 76.2, p < .001$	$F(3, 9) = .5, p = .2$
CV_1	$F(1, 9) = 13.5, p < .01$	$F(3, 9) = 2.3, p = .2$
CV_2	$F(1, 9) = 59.6, p < .001$	$F(3, 9) = 3.4, p = .08$

4. Conclusions et discussion

Les résultats de cette étude montrent que la focalisation et les gestes manuels sont coordonnés en ce sens que la focalisation "attire" le geste manuel. Les apex des gestes sont en effet soit inclus dans l'élément focalisé soit très proches. Ceci était prévisible pour le geste de pointage puisque les pointages manuel et vocal avaient le même objet qui était désigné au niveau vocal par la focalisation et au niveau gestuel par le pointage. Nous retrouvons donc bien là, dans la lignée des résultats de Rochet-Capellan et collègues [12], que la partie du geste manuel qui montre (index étendu vers la cible) et la partie de la parole qui montre (focalisation) se chevauchent. On trouve de plus que la coordination se fait essentiellement en alignant l'apex du geste avec une cible plutôt articulatoire qu'acoustique (toujours de façon cohérente avec Rochet-Capellan *et al.* [12]). L'attraction du geste manuel par la focalisation était également prévisible pour le geste de battement puisque plusieurs études ont déjà montré que le geste de battement était lié à l'emphase dans le discours [5]. Par contre, nous ne nous attendions pas à un tel résultat pour le geste de contrôle pour lequel nous pensions qu'il n'y aurait aucune coordination particulière entre le geste manuel et la focalisation. Il est possible que le geste de contrôle ait été mal choisi en ce sens qu'il est peut-être trop proche d'un geste de pointage (extension du bras et de l'index nécessaires pour aller appuyer sur le bouton). L'analyse des écart-types montre cependant que la coordination est beaucoup plus stricte pour le pointage ce qui va dans le sens de nos prévisions. Il apparaît donc que le lien fonctionnel entre geste et parole a une grande influence sur leur coordination. C'est en effet pour le geste de pointage que ce lien est le plus fort et la coordination la moins stricte est observée pour le geste de contrôle qui est un geste non-communicatif.

Les résultats montrent aussi que la production d'un geste manuel — quelque soit son type — n'a aucun effet sur les corrélats acoustiques et articulatoires de la focalisation prosodique. Il n'y a aucune différence entre la condition parole seule et les conditions parole+geste et aucune différence non plus pour les différents types de gestes. Ces résultats ne sont pas

dans la continuité de ceux de Krahmer & Swerts [13] qui avaient trouvé que la production d'un geste de battement avait apparemment pour effet d'augmenter l'activité musculaire liée à l'articulation. En fait, il est possible que les résultats de ces auteurs soient un artefact de leur protocole expérimental. Les participants devaient en effet parfois produire un geste de battement sur un mot différent du mot accentué (condition d'incongruence). Ce type de production n'est pas naturel et il est possible que les locuteurs aient tout simplement eu tendance à produire un accent aussi sur le mot concomitant à leur geste donnant ainsi l'impression d'une augmentation de certains paramètres acoustiques en fait liés à la production pure et simple d'un accent qui n'existait pas dans la condition sans geste.

Notons enfin que pendant le déroulement des expériences, nous avons pu constater que bien que les gestes de battement soient produits très fréquemment dans la communication parlée, il est très difficile de les faire produire à un locuteur "sur commande". Il serait ainsi crucial de tenter d'effectuer ces mesures dans des conditions plus naturelles.

Références

- [1] D. McNeill, *Hand and Mind : What Gestures Reveal about Thought*. University Of Chicago Press, 1992.
- [2] A. Kendon, *Gesture : Visible Action as Utterance*. Cambridge University Press, October 2004.
- [3] D. Bolinger, "Intonation and gesture," *American Speech*, vol. 58, pp. 156–174, 1983.
- [4] S. Nobe, "Representational gestures, cognitive rhythms, and acoustic aspects of speech : A network/threshold model of gesture production," Ph.D. dissertation, The Faculty of the Division of the Social Sciences, 1996.
- [5] E. McClave, "Pitch and manual gestures," *Journal of Psycholinguistic Research*, vol. 27, no. 1, pp. 69–89, 1998.
- [6] J. Boyer, A. Di Cristo, and I. Guaïtella, "Rôle de la voix et des gestes dans la focalisation," in *Oralité et gestualité. Interaction et comportements multimodaux dans la communication*, C. Cavé, I. Guaïtella, and S. Santi, Eds. L'Harmattan, 2001, pp. 459–463.
- [7] L. Pietrosemoli, E. Mora, and M. A. Blondet, "Synchronisation des mouvements des mains et de la ligne de fréquence fondamentale en espagnol parlé," in *Oralité et gestualité. Interaction et comportements multimodaux dans la communication*, C. Cavé, I. Guaïtella, and S. Santi, Eds. L'Harmattan, 2001, pp. 492–495.
- [8] D. P. Loehr, "Gesture and intonation," Ph.D. dissertation, Faculty of the Graduate School of Arts and Sciences of Georgetown University, 2004.
- [9] S. Duncan, "Gesture and speech prosody in relation to structural and affective dimensions of natural discourse," in *GESPIN - Gesture & Speech in Interaction*, 2009.
- [10] S. Kita and A. Özyürek, "What does cross-linguistic variation in semantic coordination of speech and gesture reveal? : Evidence for an interface representation of spatial thinking and speaking," *Journal of Memory and Language*, vol. 48, no. 1, pp. 16–32, January 2003.
- [11] J. P. de Ruiter, "Gesture and speech production," Ph.D. dissertation, Catholic University of Nijmegen, Netherlands, 1998.
- [12] A. Rochet-Capellan, R. Laboissière, A. Galvan, and J.-L. Schwartz, "The speech focus position effect on jaw-finger coordination in a pointing task," *Journal of Speech, Language, and Hearing Research*, vol. 51, pp. 1507–1521, December 2008.
- [13] E. Krahmer and M. Swerts, "The effects of visual beats on prosodic prominence : Acoustic analyses, auditory perception and visual perception," *Journal of Memory and Language*, vol. 57, no. 3, pp. 396–414, 2007.
- [14] M. Dohen and H. Loevenbruck, "Identification des corrélats visibles de la focalisation contrastive en français," in *Proceedings of XXVe Journées d'Etudes sur la Parole*, April 19–22 2004, pp. 185–188.
- [15] P. Boersma and D. Weenink, "Praat : doing phonetics by computer," 1995–2009. [Online]. Available : www.praat.org