

Encodage de la distance et coopération parole/geste : étude développementale du pointage multimodal

Chloe Gonseth¹ Coriandre Vilain¹ Anne Vilain¹

(1) Gipsa-Lab, 11, rue des Mathématiques, Grenoble Campus BP46 38402 St Martin D'Hères Cedex
chloe.gonseth@gipsa-lab.grenoble-inp.fr,
coriandre.vilain@gipsa-lab.grenoble-inp.fr,
anne.vilain@gipsa-lab.grenoble-inp.fr

RÉSUMÉ

Cette étude expérimentale s'intéresse, via l'utilisation du processus déictique, à l'interaction des systèmes parole/geste, et ce au cours du développement langagier. Les participants, adultes et enfants, avaient pour tâche de désigner, par un geste et/ou un mot de pointage, un objet cible, pouvant être proche ou lointain. L'utilisation conjointe ou indépendante des deux modalités déictiques, parole et geste, nous a permis de mettre en évidence une coopération entre ces deux systèmes, qui sont utilisés de manière complémentaire plutôt que redondante. Nos résultats sont également en faveur d'un encodage de la distance de l'objet pointé, non seulement par les indices articulatoires de la parole, mais aussi par les indices cinématiques du geste manuel.

ABSTRACT

Distance encoding and speech/gesture cooperation : a developmental study on multimodal pointing

The aim of this paper is to characterize, through the deictic process use, the speech/gesture interaction during language development. Participants, both adults and children, had to designate with a deictic word and/or a deictic gesture a target, which could be either close or distant. The use of two vs one modality (speech and gesture) allowed us to attest a cooperation between the two systems, which take place in a complementary rather than redundant way. Furthermore, our results are attesting that distance can be encoded not only in articulatory cues of vocal pointing, but also in kinematic cues of manual pointing.

MOTS-CLÉS : Interaction parole/geste ; Pointage déictique ; Encodage de la distance ; Développement.

KEYWORDS: Speech/gesture interaction ; Deictic pointing ; Distance encoding ; Development.

1 Introduction

Cette étude, basée sur le processus particulier qu'est la deixis spatiale, s'intéresse au rôle du couple parole/geste au cours du développement de la communication langagière.

La deixis est une capacité qui nous permet, dès le plus jeune âge, d'attirer l'attention de notre interlocuteur vers un objet ou évènement particulier, et de partager ainsi diverses informations (localisation spatiale notamment). Sa particularité réside en sa multimodalité, puisqu'au-delà de la seule utilisation du système vocal ou du système gestuel, on trouve bien souvent une com-

binasion des deux. Ainsi, les gestes déictiques s'accompagnent souvent d'un terme déictique, tel que « *here* » ou « *that* » en anglais (Kendon, 2004). Le pointage déictique, universel et omniprésent dans les interactions quotidiennes, peut être considéré comme l'un des précurseurs de la communication référentielle, à la base du développement phylogénétique de la communication langagière humaine (Arbib, 2005).

La présente étude se propose tout d'abord de caractériser, d'un point de vue ontogénétique, l'interaction parole/geste, en montrant qu'il s'agit d'une coopération entre deux modalités dépendantes l'une de l'autre. D'autre part, nous souhaitons mettre en évidence un encodage spécifique de la distance des objets pointés dont la fonction, déictique et donc communicative, serait de permettre des interactions optimales.

La distance d'un objet pointé, par rapport à la position des interlocuteurs, est en effet une caractéristique spatiale que nous pensons être encodée à la fois par les caractéristiques cinématiques des gestes manuels et par les caractéristiques articulatoires et phonétiques des termes utilisés. Cette hypothèse se base notamment sur les travaux de Bonfiglioli (Bonfiglioli *et al.*, 2009) sur l'interaction entre l'encodage lexical de la distance et la production de gestes de grasping. L'encodage lexical de la distance se fait notamment par l'utilisation de termes déictiques distaux et proximaux. Les premiers sont utilisés de préférence pour désigner des objets éloignés (comme par exemple « là »), tandis que les seconds s'adressent plutôt à des objets proches (« ici »). Certaines études typologiques ont mis en évidence une relation spécifique entre la distance encodée (i.e. proche ou lointaine) et les traits phonologiques des termes déictiques appropriés (i.e. proximaux ou distaux). Diessel (Diessel, 1999) a notamment montré une tendance intéressante dans les langues du monde : les versions distales contiennent généralement des voyelles ouvertes, contrairement aux versions proximales, qui contiennent davantage de voyelles fermées. Cette tendance au niveau lexical peut-elle être considérée comme un exemple de phonosymbolisme, i.e. un rapport non arbitraire et motivé entre phonétique et sémantique ? Notre hypothèse ici est que ce phénomène peut s'expliquer par un comportement purement moteur. Autrement dit, les locuteurs auraient tendance à ouvrir plus grand la bouche lorsqu'ils désignent un objet éloigné.

De précédents résultats (Gonseth *et al.*, 2010, 2012) ont montré que la distance de l'objet pointé, mais également les modalités utilisées pour désigner cet objet, influencent certains paramètres acoustiques et articulatoires de la parole. Le premier formant F1 ainsi que l'ouverture des lèvres sont en effet plus élevés lorsque le locuteur désigne un objet éloigné, et cette dernière l'est également lorsqu'une seule modalité est disponible pour le faire. Une telle interaction entre parole, geste et langage chez l'adulte pose la question de son développement chez l'enfant. Le pointage déictique, impliqué dans la phylogénèse de la communication référentielle apparaît également crucial du point de vue ontogénétique. Il est en effet le premier outil utilisé par les enfants pour désigner ce qui les entoure (Kita, 2003). Son utilisation est corrélée avec d'importantes étapes dans l'acquisition de productions verbales et de compétences perceptives (Ozcaliskan et Goldin-Meadow, 2005), et est ainsi bénéfique pour le développement du langage et des compétences communicatives (Butcher et Goldin-Meadow, 2000). Nous avons donc choisi de poursuivre nos travaux sur la communication multimodale de l'adulte par son étude chez l'enfant. Bien que les gestes soient extrêmement présents dans les deux populations, leur utilisation va dépendre du niveau d'expertise du locuteur. Ainsi, l'enfant utilise la modalité gestuelle comme entrée dans la communication linguistique alors que l'adulte attribue aux gestes des fonctions beaucoup plus complexes (e.g. mettre en valeur certains éléments du discours). Nous pouvons donc nous attendre à un traitement cognitif différent selon la population étudiée, et à une évolution de la coordination spatiale au cours des premières années du développement

langagier. Ceci nous permettra d'établir un agenda développemental du pointage multimodal.

2 Méthode

L'objectif de l'étude présentée ici est d'optimiser et simplifier le protocole précédemment mené chez l'adulte (Gonseth *et al.*, 2010, 2012) afin d'obtenir de nouvelles données chez l'enfant.

2.1 Participants

Vingt-neuf adultes (âgés de 18 à 36 ans) et onze enfants (âgés de six à 12 ans, voir tableau 1 ci-dessous) naïfs ont participé volontairement à cette étude.

Participants	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Age (années)	11	11	6	6	12	9	9	6	7	6	7

TABLE 1 – Age des participants - Enfants

2.2 Procédure

L'expérience a lieu en chambre sourde, où les participants sont assis face à deux diodes lumineuses (LED) situées à 140cm (D1) et 425cm (D2). Ces deux LEDs se trouvent donc dans l'espace extra-personnel du participant, c'est-à-dire hors d'atteinte. Toutefois, D1 est placée dans l'espace extra-personnel proche, alors que D2 est placée dans l'espace extra-personnel lointain. Notons que les LEDs sont alignées par rapport au participant, afin que l'angle de son bras ne puisse à lui seul indiquer la diode pointée, un geste spécifique pouvant alors s'avérer nécessaire. En réponse à la question de l'expérimentateur « C'est où ? », le participant doit désigner la LED allumée, selon les trois conditions suivantes (chaque condition comporte 20 essais, l'ensemble étant présenté de manière aléatoire) :

- Parole seule : Nommer la diode allumée avec le déictique « là » inséré dans une phrase porteuse (i.e. précédé de « c'est »).
- Geste seul : Pointer la diode allumée avec l'index.
- Parole+Geste : Simultanément pointer et nommer la diode allumée.

Nous attendons des productions vocales et manuelles renforcées en condition unimodale (i.e. Parole seule ou Geste seul) par rapport à la condition bimodale (Parole+Geste), mais également lorsque les participants désignent la diode située en D2.

2.3 Enregistrements & mesures

2.3.1 Données articulatoires & manuelles

Les mouvements oro-faciaux et ceux de l'index sont enregistrés par le système de capture de mouvements Optotrak 3020. Trois émetteurs de référence sont placés sur le front du participant,

deux émetteurs sur les lèvres (supérieure et inférieure), et un émetteur sur l'index. Concernant la modalité vocale, l'ouverture des lèvres est ainsi mesurée : OL = lèvre du bas - lèvre du haut, en mm. Concernant la modalité gestuelle, la durée du plateau (i.e. durée de maintien en position de pointage, en s), le pic de vitesse de l'aller du geste (en mm/s), ainsi que l'amplitude (en mm) sont mesurés.

2.3.2 Données acoustiques

Les productions orales des participants sont également enregistrées par microphone AKG C1000S. En complément des données articulatoires, nous avons choisi d'étudier un paramètre acoustique particulier, supposé corrélé à l'ouverture des lèvres. Ainsi, F1 (en Hz) est calculé au milieu de la voyelle [a] de chaque production du déictique /la/.

3 Résultats

Les items contenant des erreurs de production (anticipation, retard, oubli) sont exclus de l'analyse statistique. Pour cette raison, deux adultes et trois enfants ne seront pas inclus dans l'analyse finale. Une analyse de la variance à mesures répétées est effectuée pour chaque groupe (adultes et enfants). Elle porte sur les médianes de chaque type de mesure (articulatoire, acoustique et manuelle), avec un seuil de significativité fixé à $p < 0.05$ et les deux facteurs d'intérêt suivants :

- Trois conditions C : Parole - Geste - Parole+Geste
- Deux distances D : 140 cm - 425 cm

A noter que les barres d'erreur standard sont représentées sur les graphiques, sections 3.1 3.2.

3.1 Données articulatoires & acoustiques

3.1.1 Ouverture des lèvres

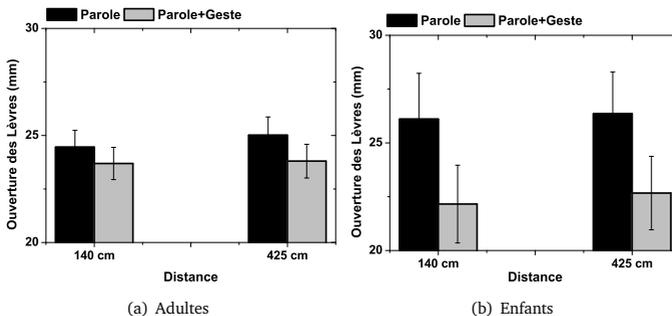


FIGURE 1 – Ouverture des lèvres en fonction de la condition et de la distance

La figure 1(a) illustre les variations articulatoires présentées par les adultes, en fonction de nos deux facteurs d'intérêt. Les résultats montrent un effet significatif de la condition sur l'ouverture des lèvres ($F(1, 24)=17.06, p<0.05$), cette dernière étant plus élevée en condition unimodale (Parole seule) que bimodale (Parole+Geste). La distance influence également l'ouverture des lèvres, puisqu'indépendamment de la condition, celle-ci est significativement plus élevée pour la distance la plus grande ($F(1, 24)=8.87, p<0.05$).

La figure 1(b) concerne quant à elle les variations articulatoires présentées par les enfants. Les résultats sur l'effet de la condition sont similaires à ceux des adultes ($F(1, 7)=19.89, p<0.05$) : l'ouverture des lèvres est plus grande en situation unimodale. En revanche, celle-ci ne dépend pas de la distance désignée ($F(1, 7)=1.46, p=0.27, ns$).

3.1.2 F1

La figure 2(a) illustre les variations acoustiques présentées par les adultes. Les valeurs de F1 sont significativement plus élevées en condition unimodale ($F(1, 26)=28.03, p<0.05$) mais ne dépendent pas de la distance de l'objet pointé ($F(1, 26)=0.84, p=0.36, ns$).

La figure 2(b) concerne les variations acoustiques présentées par les enfants. Les valeurs de F1 ne sont influencées ni par la condition ($F(1, 7)=0.14, p=0.72, ns$), ni par la distance ($F(1, 7)=1.54, p=0.25, ns$).

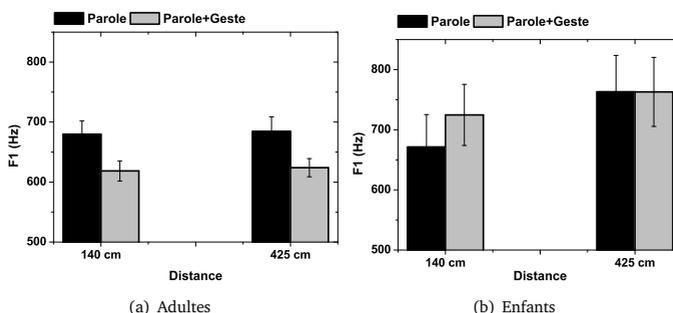


FIGURE 2 – F1 en fonction de la condition et de la distance

3.2 Données manuelles

Les figures 3(a), 3(b) et 3(c) illustrent les variations cinématiques du geste de pointage chez l'adulte. Nous observons que la durée du plateau est significativement plus élevée en condition unimodale ($F(1, 24)=32.60, p<0.05$), alors que la vitesse est significativement plus élevée en condition bimodale ($F(1, 24)=16.51, p<0.05$). L'amplitude en revanche ne varie pas en fonction de la condition ($F(1,24)=0.93, p=0.34, ns$). D'autre part, la distance de l'objet pointé a quant à elle un effet significatif sur chaque paramètre (Plateau $F(1, 24)=19.84, p<0.05$; Vitesse $F(1,$

24)=16.89, $p<0.05$; Amplitude $F(1, 24)=13.47$, $p<0.05$). Un geste de pointage vers un objet distant sera plus long, plus rapide et plus ample.

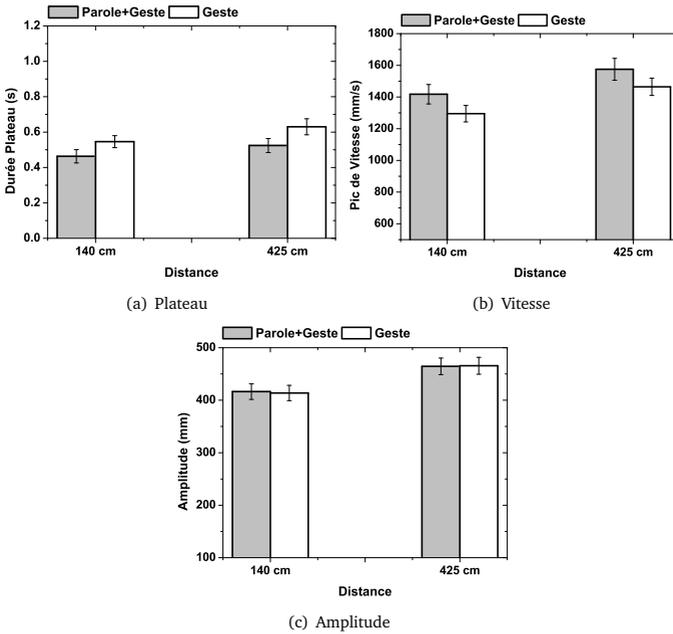


FIGURE 3 – Cinématique du geste en fonction de la distance et de la condition - Adultes

Les figures 4(a),4(b) et 4(c) présentent les variations cinématiques du geste de pointage chez l'enfant. Tout comme chez l'adulte, la durée du plateau est plus élevée en condition unimodale ($F(1,7)=17.5$, $p<0.05$) et l'amplitude du geste est indépendante de la condition ($F(1,7)=2.77$, $p=0.14$, ns). Contrairement aux adultes, les enfants n'augmentent pas la vitesse de leurs pointages en condition bimodale ($F(1,7)=1.26$, $p=0.3$, ns). Concernant l'effet de la distance, des résultats similaires à ceux des adultes sont observés : la distance influence chacun des trois paramètres, qui augmentent pour la distance la plus éloignée (Plateau $F(1,7)=5.35$, $p<0.05$; Vitesse $F(1,7)=27.28$, $p<0.05$; Amplitude $F(1,7)=18.82$, $p<0.05$).

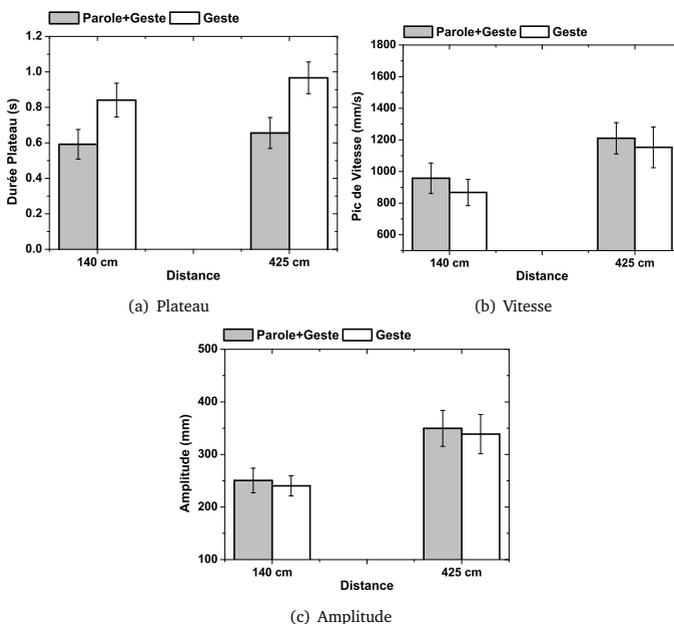


FIGURE 4 – Cinématique du geste en fonction de la distance et de la condition - Enfants

4 Discussion & conclusion

Les résultats de cette étude sont en faveur d'une coopération entre les systèmes vocal et gestuel, qui sont utilisés de manière complémentaire plutôt que redondante, l'information étant distribuée en fonction des modalités disponibles pour la transmettre. Nous remarquons chez l'adulte un effet significatif de la condition (uni vs bimodale) sur les valeurs articulatoires, acoustiques (effet qui n'était pas observé dans nos précédents travaux, cités section 1) et cinématiques du pointage de l'index. Les enfants présentent des résultats similaires à ceux de l'adulte au niveau articulatoire et cinématique. Cela suppose que les participants compensent l'absence de pointage manuel par un pointage vocal renforcé, et inversement (le pointage manuel dépendant lui aussi de la présence du pointage vocal).

D'autre part, nos résultats sont en faveur d'un encodage de la distance dans les deux modalités de pointage. Chez l'adulte uniquement, cet encodage se traduit au niveau articulatoire par une ouverture des lèvres plus grande pour la distance la plus éloignée. Cet effet, observé à l'intérieur d'une même catégorie vocalique, suggère que la tendance des langues du monde à associer voyelles fermées et déictiques proximaux, voyelles ouvertes et déictiques distaux pour-

raient provenir d'un corrélat purement moteur plutôt que lexical. Notons toutefois que F1 tend à augmenter avec la distance, mais, contrairement à ce que nous avons observé dans nos précédents travaux, cités section 1, cette tendance n'est pas significative. Au niveau cinématique, l'encodage de la distance se traduit, chez les adultes et les enfants, par un geste plus long, plus rapide et plus ample, lorsque dirigé vers un objet distant. L'encodage de la distance semble donc effectif chez l'adulte dans les deux modalités de pointage, alors qu'il ne l'est chez l'enfant qu'à travers la modalité gestuelle. L'encodage vocal de la distance serait donc un outil pragmatique plus complexe, au contraire de la coopération parole/geste, qui serait un outil de base de la communication langagière.

La coopération parole/geste ainsi que l'encodage de la distance pourraient améliorer l'interaction, en rendant la communication plus économique et efficace. Mais ces deux outils sont-ils des caractéristiques fondamentales de la communication langagière, ou des fonctions cognitives plus élaborées ? Il est essentiel de poursuivre cette étude pour établir des profils développementaux précis et par groupe d'âge, afin de pouvoir répondre à ces questions.

Remerciements

Nous remercions chaleureusement les sujets adultes et enfants pour leur participation. Un grand merci à Benjamin Roustan, Marion Dohen, Hélène Loevenbruck, Marc Sato et Jean-Luc Schwartz, pour leurs précieux conseils et discussions enrichissantes.

Références

- ARBIB, M. A. (2005). From monkey-like action recognition to human language : An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28:105–167.
- BONFIGLIOLI, C., FINOCCHIARO, C., GESIERICH, B., ROSITANI, F. et MASSIMO, V. (2009). A kinematic approach to the conceptual representations of this and that. *Cognition*, 111:270–274.
- BUTCHER, C. et GOLDIN-MEADOW, S. (2000). *Language and Gesture*, chapitre Gesture and the transition from one- to two-word speech : When hand and mouth come together, pages 235–257. Cambridge University Press, New York.
- DIESSEL, H. (1999). *Demonstratives : Form, Function, and Grammaticalization*. John Benjamins, Amsterdam.
- GONSETH, C., VILAIN, A. et VILAIN, C. (2010). Speech/gesture relationship in deictic pointing. In *Proceedings of the 12th Conference on Laboratory Phonology (LabPhon12)*, Albuquerque, New-Mexico.
- GONSETH, C., VILAIN, A. et VILAIN, C. (2012). Deictic pointing : How do speech and gesture cooperate to encode distance information. *Submitted*.
- KENDON, A. (2004). *Gesture : Visible Action as Utterance*. Cambridge University Press, Cambridge.
- KITA, S. (2003). *Pointing, Where Language, Culture, and Cognition Meet*. Psychology Press.
- OZCALISKAN, S. et GOLDIN-MEADOW, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96(3):101–113.