

# Dénomination d'image versus détection interne de phonème : deux méthodes pour étudier la planification de la production de parole

Pierre Hallé<sup>1,2,3</sup> Laura Manoiloff<sup>4</sup> Juan Segui<sup>2</sup>

(1) Laboratoire de Phonétique et Phonologie, 75005 Paris, France

(2) Laboratoire Mémoire et Cognition, 92100 Boulogne-Billancourt, France

(3) Haskins laboratories, CT 06511 New Haven, USA

(4) Laboratorio de Psicología Cognitiva, X5000 Cordoba, Argentina

pierre.halle@univ-paris3.fr, lmvmanoiloff@gmail.com,

juan.segui@parisdescartes.fr

## RESUME

---

Cette étude est motivée initialement par une question méthodologique : la validité des mesures de temps de dénomination d'image, très utilisés pour explorer les processus de planification de production de parole. Le temps de dénomination est le temps écoulé entre affichage de l'image et début acoustique de la réponse verbale. Dans cet article, nous résumons la littérature sur les inconvénients de cette mesure. Nous présentons ensuite notre étude, qui compare directement temps de dénomination d'image et temps de détection interne de phonème initial. Les participants sont hispanophones. Les noms d'image sont contrastés en fréquence lexicale et phonème initial. Les temps de réponse pour les deux mesures sont assez proches. Cependant, ceux de détection de phonème sont relativement insensibles au type de phonème initial, contrairement aux temps de dénomination. Au delà de l'avantage méthodologique de la détection interne de phonème, nos données suggèrent que celle-ci opère sur des représentations relativement abstraites.

## ABSTRACT

---

### **Picture naming versus internal phoneme monitoring: two methods for exploring speech production planning.**

A methodological issue initially motivated the present study: how valid are the measurements of picture-naming times, which are widely used to investigate the processes of speech production planning? Picture-naming time is the time elapsed from picture display to the acoustic onset of verbal response. We begin with a review of the literature pointing to the shortcomings of naming times. We then report our own study, which directly compares picture-naming times and internal monitoring times of word-initial phonemes. Response times for both measurements are quite similar. However, the phoneme detection response times are relatively insensitive to word-initial phoneme's phonetic type, contrary to picture-naming times. Beyond the methodological advantage of internal phoneme monitoring over naming, our data suggest that internal monitoring bears on rather abstract internal representations.

---

**MOTS-CLES :** production de parole, dénomination d'image, détection interne de phonème.

**KEYWORDS:** speech production planning, picture-naming, internal phoneme monitoring.

---

# 1 Introduction

La tâche de dénomination d'image a une très longue histoire (Cattell, 1885). Elle a été et reste encore très utilisée pour l'étude de la production de parole. Dans cette tâche, on présente au sujet une série d'images, une par une, que le sujet doit nommer aussi rapidement que possible. La variable dépendante est le "temps de dénomination" mesuré entre l'apparition de l'image et le début de la réponse vocale. De nombreuses études ont utilisé un dispositif appelé "clef vocale" pour effectuer cette mesure. La clef vocale se "déclenche" au début de la réponse vocale, mesurant alors un temps de réponse. Il est naturellement possible que le sujet produise une réponse "incorrecte" par exemple *chat* pour l'image d'un chien, voire un hypéronyme non attendu comme *quadrupède*, *animal*, etc. pour le même chien, voire encore qu'il ne réponde pas mais que la clef se déclenche malgré tout, par exemple sur un bruit de toux. Il est donc nécessaire de vérifier les réponses en temps réel ou bien d'enregistrer les productions vocales des sujets et vérifier après l'expérience si les mesures de la clef vocale sont acceptables ou non.

Même après ces vérifications, les valeurs données par les clefs vocales ne sont pas entièrement satisfaisantes. Rastle et Davis (2002) ont montré que les deux types existants de clef vocale (à seuil simple ou à intégration) se déclenchent systématiquement *après* le départ acoustique. Il ressort clairement de cette étude que la façon de mesurer les temps de dénomination la plus précise est d'enregistrer les réponses vocales et de repérer le départ acoustique de l'énoncé produit par le sujet — s'il s'agit d'une réponse correcte — à partir du signal acoustique et son spectrogramme.

Même en appliquant cette procédure quelque peu coûteuse (mais qui peut être assistée de façon semi-automatique par un algorithme de détection de départ acoustique), un problème subsiste lorsque l'on souhaite comparer des mots commençant par des consonnes différentes. En particulier, la variation de temps de dénomination induite par les différences entre consonnes initiales a été examinée par Rastle et al. (2005). Ces auteurs utilisent une tâche de dénomination retardée de transcriptions phonétiques (les sujets ont au moins 5 ans de formation en phonétique). Les sujets ne prononcent la transcription affichée sur un écran que lorsqu'un ton bref a été émis. Cette procédure permet de mesurer des temps de dénomination une fois que la phase de préparation de la production est achevée : seules des différences liées à l'articulation des réponses verbales peuvent influencer les temps. Rastle et al. observent que des différences phonétiques entre phonèmes initiaux induisent une variation substantielle et suggèrent aux chercheurs qui utilisent les temps de dénomination de se limiter à la comparaison de mots cibles partageant le même phonème initial. Nous avons repris leurs données dans la figure 1 qui montre clairement que les temps les plus longs sont systématiquement associés aux occlusives, en particulier les /b, d, g/ anglais réalisés [p, t, k]

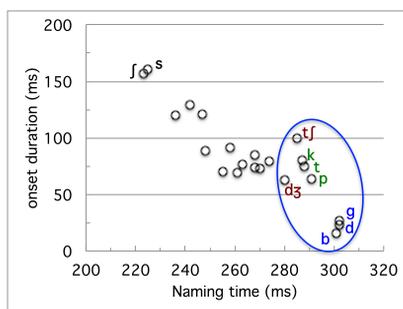


FIGURE 1 : temps de dénomination x durée acoustique du phonème initial, par phonème.

Nous nous plaçons ici dans le cadre du modèle de Levelt (e.g., [Levelt et al., 1999](#)). Dans ce modèle, l'accès lexical, par exemple au nom d'une image, fournit une représentation phonologique abstraite consistant d'une part en une séquence de phonèmes et d'autre part en un cadre métrique et prosodique. Ces informations sont combinées, éventuellement en fonction des mots adjacents, par des opérations d'encodage phonologique dont le résultat est une spécification phonologique complète intégrant la métrique, et en particulier la syllabation. Levelt considère cette spécification syllabée comme le point de départ de la construction d'un programme articuloire (qu'il appelle 'phonetic plan') constitué de 'gestural scores' syllabiques. Dans ce cadre de travail, le temps de réponse que l'on voudrait idéalement mesurer est celui de la sortie de l'encodage phonologique, point de départ pour lancer l'articulation. Le plus proche de ce temps idéal est le départ *articuloire* de la réponse vocale, modulo un décalage que l'on peut supposer relativement stable, bien qu'il varie sans doute en fonction des différents ensembles d'articulateurs engagés dans l'articulation physique et donc des différents paramètres physiologiques d'inertie et de contrôle musculaire qui leur sont associés.

Les données de [Rastle et al. \(2005\)](#) montrent que le temps de dénomination s'éloigne nettement de cet idéal, même lorsqu'il est mesuré avec précision, entre le début de l'affichage de l'image et le début acoustique de la réponse vocale. C'est en particulier le cas lorsque le mot produit commence par une occlusive non-voisée. Pour ces occlusives, le départ articuloire est le début du geste d'occlusion. Il précède le début acoustique qui ne peut-être que le relâchement de l'occlusion. On ne peut observer la trace acoustique du début de l'occlusion que dans le cas des occlusives pré-voisées : c'est approximativement le début du pré-voisement. Il n'est donc pas étonnant que [Rastle et al. \(2005\)](#) trouvent des temps de dénomination plus longs pour les occlusives —y compris les affriquées— que les sonantes et les fricatives, puisque toutes les occlusives en position initiale sont non-voisées phonétiquement en anglais. C'est bien ce que montre la figure 1.

Cependant, mesurer le départ articuloire requiert un appareillage spécialisé et n'est de toute façon pas une chose facile à mettre en œuvre. [Schaeffler et al. \(2014\)](#) ont comparé les temps de réponse de dénomination d'image mesurés à partir de l'articulation (enregistrements ultra-son de la langue et vidéo des lèvres) vs. à partir du signal acoustique. Ils trouvent que le départ articuloire précède le départ acoustique d'environ 120 à 180 ms. Autrement dit, la parole audible est précédée par des mouvements silencieux des articulateurs. [Schaeffler et al. \(2014\)](#) trouvent aussi que le lieu (labial, lingual) et la manière (voyelle, occlusive, fricative) d'articulation du phonème initial n'ont pas d'effet sur les temps mesurés pour les départs articuloires. Donc, en particulier, pas de différence entre occlusives et fricatives. Ces mesures se rapprochent donc de la mesure idéale mentionnée plus haut mais au prix d'une lourdeur technique prohibitive pour les études de production de parole.

Dans cette étude, exploratoire par sa dimension, nous comparons temps de dénomination d'image et temps de détection interne du phonème initial du nom de l'image avec le même matériel expérimental, c'est à dire les mêmes images. Nous tentons d'abord de répondre à la question suivante : Quelle est la sensibilité des deux mesures de temps de réponse aux variations liées aux différents types de phonème initial ? L'étude de [Rastle et al. \(2005\)](#) suggère que, toutes choses égales par ailleurs, les temps de dénomination mesurés à partir du signal acoustique sont plus longs pour les occlusives que les autres types de phonème. L'étude de [Schaeffler et al. \(2014\)](#) suggère que le type de phonème initial, par exemple occlusive vs. fricative, n'a pas d'effet sur les temps de dénomination mesurés à partir de l'articulation. En est-il de même pour les temps de détection interne de phonème initial ? Un objectif secondaire de cette étude est de comparer les effets bien connus de fréquence lexicale en production ([Oldfield et Wingfield, 1965](#)) sur les temps de dénomination et les temps de détection interne. L'effet de fréquence lexicale sur les temps de

détection interne de phonème initial a été mis en évidence par [Manoiloff et al. \(2013\)](#). Cependant, ces effets n'ont encore jamais été comparés avec ceux obtenus pour les temps de dénomination : sont-ils comparables pour les deux mesures ? Ou bien l'une des deux est-elle plus sensible que l'autre ? Pour répondre à cette question ainsi qu'à la question principale sur les effets du type de phonème initial, nous avons repris le matériel utilisé par [Manoiloff et al. \(2013\)](#). Ceci permet en même temps de vérifier si les effets de fréquence lexicale en détection interne de phonème initial peuvent ou non être répliqués.

## 2 Expériences : dénomination vs. détection interne de phonème

Nous avons testé deux groupes d'une trentaine de sujets argentins hispanophones sur les mêmes images. Pour un groupe, les sujets devaient dénommer les images présentées. Pour l'autre groupe, les sujets devaient détecter des phonèmes pré-spécifiés dans le nom des mêmes images.

### 2.1 Méthode

#### – Participants

Soixante-et-un étudiants en Psychologie à l'Université Nationale de Cordoba (Argentine), âgés entre 18 et 26 ans, ont participé volontairement aux expériences. Trente ont participé à l'expérience de détection de phonème et 31 autres à l'expérience de dénomination d'image. Tous étaient locuteurs natifs de l'espagnol parlé en Argentine. Aucun d'entre eux ne souffrait de déficit de vision ou de trouble du langage.

#### – Matériel

Nous avons repris le matériel utilisé par [Manoiloff et al. \(2013\)](#). Ce matériel consiste en 30 dessins expérimentaux en noir et blanc, sélectionnés parmi les 400 dessins de l'étude de [Cycowicz et al. \(1997\)](#), qui inclut les 260 dessins de [Snodgrass et Vanderwart \(1980\)](#). Quinze dessins ont un nom de haute fréquence et 15 un nom de basse fréquence, selon les tables de fréquence établies par [Alameda et Cuetos \(1995\)](#) et selon le jugement subjectif de trente sujets n'ayant pas participé à l'étude. La table 1 montre les caractéristiques de ces trente noms d'image. Le pourcentage d'accord sur le nom des images et leur complexité visuelle sont tirés des normes de [Manoiloff et al. \(2010\)](#). Il est à noter que les pourcentages d'accord sur le nom sont très élevés. À ces 30 items expérimentaux s'ajoutent 227 items de remplissage pour la phase de test et 18 ou 20 items pour une phase d'entraînement familiarisant les sujets à l'une des deux tâches. Ces items supplémentaires étaient extraits du matériel utilisé par [Cycowicz et al. \(1997\)](#) et [Manoiloff et al. \(2010\)](#).

| caractéristiques                                   | noms HF | noms BF |
|--|---------|---------|
| fréquence des tables de Alameda et Cuetos (o.p.m.) | 138.1   | 5.3     |
| fréquence subjective (échelle 1-5)                 | 4.1     | 1.6     |
| nombre de syllabes                                 | 2.3     | 2.5     |
| nombre de phonèmes                                 | 5.1     | 5.8     |
| pourcentage d'accord sur le nom                    | 98.8%   | 96.7%   |
| complexité visuelle (échelle 1-5)                  | 2.6     | 3.0     |

TABLE 1 : Caractéristiques du matériel : noms de haute vs. basse fréquence d'usage (HF vs. BF).

Bien que ce matériel n’ait pas été conçu pour tester l’effet possible du type de phonème initial sur la dénomination ou la détection interne, il se trouve que les types principaux qu’il est intéressant de tester sont bien représentés, même si de façon non-homogène. Parmi les 30 noms d’image, 17 commencent par une occlusive sourde et 12 par une consonne non-occlusive. Le mot *gato* “chat” est le seul représentant de la catégorie occlusive sonore : pour cette raison et pour d’autres raisons qui sont exposées par la suite, nous n’avons pas retenu les données pour cet item. La table 2 montre les différentes consonnes initiales des 30 noms d’image utilisés.

| Consonne | l | r | m | f | g | p  | t |
|----------|---|---|---|---|---|----|---|
| Compte   | 2 | 4 | 3 | 3 | 1 | 12 | 5 |

TABLE 2 : Distribution des consonnes initiales pour les 30 noms d’image.

#### – Design

Pour l’expérience de détection de phonème, les items étaient présentés en 7 blocs correspondant chacun à l’un des 7 phonèmes à détecter. Ces blocs comprenaient entre 25 et 72 items : les items test (entre 1 et 12 : cf. Table 2), des items de remplissage avec le phonème initial des items test (entre 0 et 4), et des items de remplissage avec un phonème initial différent (entre 20 et 60 de sorte à homogénéiser la proportion d’items test). L’ordre de présentation des blocs, et celui des items à l’intérieur des blocs, était pseudo-aléatoire et différent pour chaque sujet.

Pour l’expérience de dénomination, les mêmes items (au total,  $257 = 30 + 227$ ) étaient présentés aux sujets dans un ordre pseudo-aléatoire différent pour chaque sujet.

#### – Procédure

Pour la détection de phonème, l’expérimentateur expliquait aux sujets qu’on allait leur présenter des jeux d’images, une image à chaque essai, et qu’ils devaient répondre le plus rapidement possible en appuyant sur le bouton-réponse aux images dont le nom commençait par le phonème-cible spécifié pour le jeu en cours. Pour chacun des sept blocs (i.e., des sept “jeux”) d’images, l’expérimentateur indiquait oralement (en espagnol) le phonème à détecter par des exemples (e.g., “maintenant vous devez répondre aux images dont le nom commence par le son /t/ comme dans *tambour*, *taureau*, etc.”). Il était bien précisé au sujet qu’il devait répondre à un son et pas à une lettre. L’expérience proprement dite (la phase de test) était précédée par une phase d’entraînement comprenant six blocs de trois essais, chacun associé à un phonème-cible non utilisé dans la phase de test.

Pour la dénomination, il était expliqué aux sujets qu’ils devaient dénommer l’image présentée à chaque essai aussi rapidement que possible. Les sujets portaient un micro-casque pour éviter une trop grande variabilité de l’intensité sonore. Les réponses vocales étaient directement numérisées et enregistrées (le temps ‘zéro’ correspondant à l’apparition de l’image à dénommer) pour repérage ultérieur du départ acoustique des mots produits par les sujets. Le dispositif comprenait aussi une estimation logicielle du temps de réponse (à partir du signal acoustique enregistré). Nous n’avons pas analysé les temps donnés par cette estimation. La phase de test était précédée par une phase d’entraînement comportant 20 images non utilisées dans la phase de test.

Pour les deux expériences, les images étaient présentées sur écran d’ordinateur en noir sur fond blanc dans un carré de 7 cm de côté au centre de l’écran. Chaque image restait affichée jusqu’à la réponse du sujet (indiquée par le déclenchement du bouton réponse ou par la clef vocale) ou bien pendant deux secondes en cas d’absence de réponse. L’essai suivant était initié une seconde après la

disparition de l'image. Pour les deux expériences, les phases d'entraînement et de test étaient précédées par une phase de "familiarisation" où les 30 images expérimentales étaient présentées une par une, avec le nom attendu affiché en dessous, pendant 2 s avec un intervalle de 1 s entre deux images. Aucune réponse n'était demandée aux sujets. Cette phase était destinée à minimiser les réponses inattendues malgré le pourcentage élevé d'accord sur le nom des images (>97%). L'expérience était contrôlée avec le logiciel DMDX (Forster et Forster, 2003) et durait environ 40 minutes (détection) ou 30 minutes (dénomination).

## 2.2 Résultats

### – Données retenues

Un des 31 sujets de l'expérience de dénomination a été exclu des analyses car les réponses vocales enregistrées étaient beaucoup trop bruitées. L'item gato ("chat") a été exclu des analyses car il a été produit 11 fois sur 30 avec une fricative proche de [ɣ] au lieu de l'occlusive pré-voisée [g]. Les données manquantes de dénomination (pas de réponse avant 2 s ou réponse erronée : e.g., *ganzo* "oie" pour *pato* "canard") représentent 2.5% des 870 essais retenus (29 items x 30 sujets). Les données manquantes de détection (pas de réponse) représentent 3.9% des essais. Nous avons aussi exclu les temps de réponse supérieurs à 1500 ms. Ce qui élimine 2.5% des données de détection mais aucune donnée de dénomination.

### – Effet de fréquence lexicale

La figure 2 montre les temps de réponse (TR) de dénomination et de détection. L'effet de fréquence est manifeste pour les deux tâches. Nous avons conduit des analyses de variance par sujets (F1) et par items (F2) avec TR comme variable dépendante, la Fréquence lexicale (haute vs. basse) comme facteur inter-item et intra-sujet et la Tâche (dénomination vs. détection) comme facteur intra-item et inter-sujet. Le facteur Fréquence est très significatif globalement,  $F_1(1,58)=163.43$ ,  $p<.00001$ ;  $F_2(1,27)=25.34$ ,  $p<.00001$ , mais aussi pour chacune des deux tâches,  $p_1s<.00001$ ,  $p_2s<.001$ . Le facteur Tâche n'est pas significatif,  $F_1(1,58)=1.76$ ,  $p=.19$ ;  $F_2(1,27)=2.19$ ,  $p=.15$ . Cependant, l'interaction Tâche x Fréquence, marginalement significative par sujets,  $F_1(1,58)=3.42$ ,  $p=.069$ , suggère un effet de fréquence plus grand pour la dénomination que la détection (121 vs. 90 ms). Cette différence n'est cependant ni robuste ni numériquement impressionnante : l'effet de fréquence lexicale est de l'ordre de 100 ms pour les deux tâches, ce qui est proche des résultats de [Manoiloff et al. \(2013\)](#). De plus, les TRs de détection de phonème ne sont pas significativement plus longs, dans l'ensemble, que les TRs de dénomination.

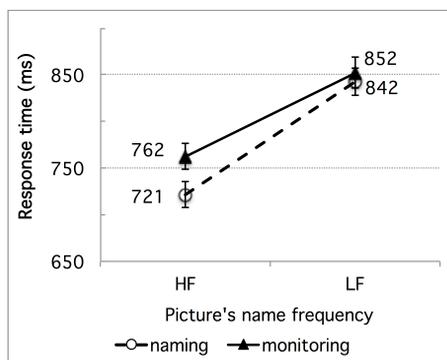


FIGURE 2 : TRs en fonction de la tâche (dénomination vs. détection) et de la fréquence.

## – Effet du type de phonème initial

Nous avons regroupé les phonèmes retenus en deux catégories : occlusives non-voisées (/p, t/) et autres (/f, m, l, r/). Comme le montre la figure 3A, les temps de réponses pour ces deux catégories sont en moyenne pratiquement les mêmes pour la détection interne, alors qu'ils sont nettement plus longs avec les occlusives que les non-occlusives pour la dénomination. Cette observation est confirmée par deux analyses de variance des temps de réponse, par sujets et par items, avec la Tâche (dénomination vs. détection) comme facteur intra-item et inter-sujet, et le type de Consonne initiale (occlusive vs. non-occlusive) comme facteur inter-item et intra-sujet. L'interaction entre ces deux facteurs n'est pas significative dans l'analyse par items mais l'est dans celle par sujets,  $F(1,58)=7.63, p<.01$ , indiquant un effet du facteur Consonne significatif pour la dénomination,  $F(1,29)=33.25, p<.00001$ ;  $F(1,27)=3.34, p=.079$  (marginal), mais pas la détection,  $F(1,29)=2.78, p=.11$ ;  $F(1,27)=1.03, p=.32$ . Ceci suggère que la tâche de détection est moins sensible que celle de dénomination aux variations du type de phonème initial du mot cible. Mais cette analyse ne tient cependant pas compte de l'effet de fréquence lexicale qui, comme le montrent la figure 2 et les statistiques correspondantes, est extrêmement robuste. Il est en principe possible que l'effet du type de phonème initial que nous trouvons (le plus clairement dans l'analyse par sujets) soit confondu en partie avec l'effet de fréquence, bien que les items expérimentaux aient été répartis à peu près équitablement dans les deux sous-ensembles haute et basse fréquence. À tout le moins, la fréquence lexicale peut perturber voire fausser la comparaison entre occlusives et non-occlusives. Des analyses de covariance, avec la fréquence comme covariable, sont donc nécessaires pour mesurer les effets possibles du type de phonème initial indépendamment de la fréquence. Nous avons conduit deux telles analyses de covariance, par sujets et par items, avec la fréquence subjective des noms d'image comme covariable. Nous trouvons que l'effet du facteur Consonne (occlusive vs. non-occlusive) est significatif pour les temps de dénomination,  $F(1,177)=17.79, p<.00001$ ;  $F(1,26)=3.26, p=.083$  (marginal), mais pas pour les temps de détection,  $F_1$  et  $F_2 < 1$ , ce qui confirme l'analyse précédente. La covariation des temps de réponse avec la fréquence subjective et le phonème initial est illustrée par la figure 3B où les six phonèmes initiaux /l, m, t, p, r, f/ sont classés par ordre décroissant de fréquence subjective. La figure suggère une relation proche de la linéarité entre TR de détection et fréquence subjective, à l'exception des items en /f/. La relation est moins linéaire pour les TRs de dénomination, systématiquement plus courts que ceux de détection pour tous les phonèmes autres que /p/ et /t/, y compris /f/.

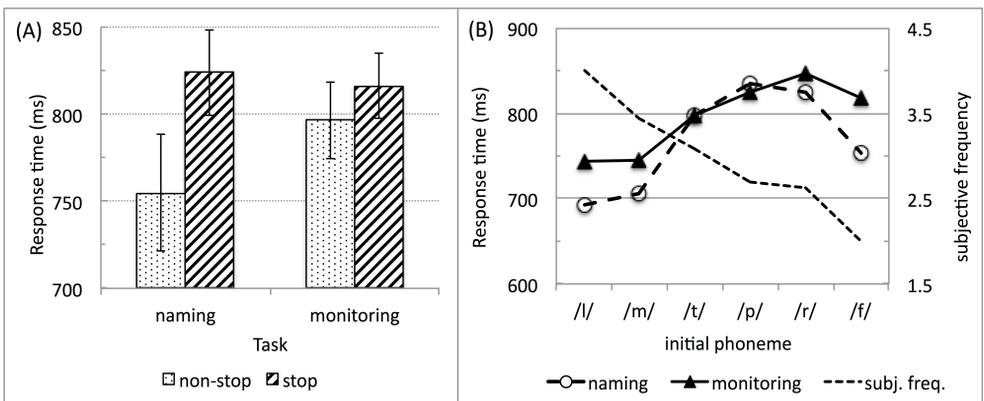


FIGURE 3 : (A) TRs en fonction de la tâche et du type de phonème initial (occlusive ou non) ; (B) TRs détaillés selon les 6 phonèmes utilisés, en ordre décroissant de fréquence subjective.

### 3 Discussion

Dans cette étude comparant temps de dénomination et temps de détection de phonème initial sur le même ensemble d'images, celui utilisé par [Manoiloff et al. \(2013\)](#), nous répliquons l'effet de fréquence lexicale ([Manoiloff et al., 2013](#)) pour les deux types de temps de réponse. L'effet est un peu plus grand pour la dénomination que pour la détection (121 vs. 90 ms), mais la différence n'est pas significative. Cependant, la question principale à laquelle cette étude tente de donner des éléments de réponse est celle de la sensibilité comparée des deux mesures de temps de réponse au type de phonème initial des noms d'image à produire, soit ouvertement (dénomination), soit intérieurement (monitoring interne). Nos données suggèrent que les temps de dénomination sont systématiquement plus longs lorsque les noms d'image commencent par une occlusive que par un autre type de consonne. Les temps de détection ne montrent pas cette différence. Ce résultat est indépendant des effets de fréquence lexicale, comme le montrent nos analyses de covariance.

Une première conséquence de nos résultats, sous réserve qu'ils soient robustes et reproductibles pour d'autres langues que l'espagnol, est un avantage méthodologique du monitoring interne sur la procédure classique de dénomination. En effet, il devient possible de comparer des ensembles de noms d'image sans qu'ils commencent nécessairement par le même type de phonème, puisque les temps de détection interne de phonème initial dans le nom d'une image semblent peu sensibles au type phonétique de phonème. En particulier, les temps de détection, contrairement aux temps de dénomination, ne sont pas plus longs, à fréquence lexicale égale, pour les occlusives que les fricatives, nasales ou liquides.

Au delà de cet avantage méthodologique en faveur du monitoring interne, nos données ont des implications pour le niveau de représentation sur lequel opère la détection interne ou plus largement, le monitoring interne en production de parole. Elles suggèrent que les représentations en jeu sont relativement abstraites. Dans le cadre du modèle de Levelt ([Levelt et al., 1999](#)), ces représentations pourraient être celles qui sont récupérées lors de l'accès aux informations lexicales, celles qui sont élaborées au cours de l'encodage phonologique, ou bien celles qui sont produites en sortie de l'encodage phonologique. La dernière possibilité est plus compatible avec la vision de l'encodage phonologique comme un module imperméable à l'inspection consciente : seule sa sortie est accessible, pas ses calculs internes. Une autre possibilité est que les phonèmes soient détectés par le système de compréhension de la parole alimenté par la sortie de l'encodage phonologique (ce que Levelt appelle la "boucle interne" 'internal loop'). Les effets de point d'unicité trouvés par [Özdemir et al. \(2007\)](#) vont dans ce sens : ces effets sont en effet typiques du traitement de la parole externe et sont liés à l'accès lexical en modalité auditive. Or, dans une situation de monitoring interne, l'accès lexical a déjà été effectué, avant l'encodage phonologique. Ce qui suggère que la détection ne se fait pas sur la sortie de l'encodeur mais se fait par la boucle interne. D'autres données récentes remettent en question la communauté des traitements de la parole externe et interne ([Gauvin et al., 2016](#)). Nos propres données ne permettent pas de trancher la question de savoir si le monitoring interne peut passer par d'autres voies que le système de compréhension/perception de la parole qui serait commun à la parole externe et à la parole interne. Elles suggèrent simplement que le monitoring interne examine des représentations abstraites indépendantes de la spécification du détail phonétique de réalisation des mots.

## Références

- ALAMEDA J., CUETOS F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de las publicaciones de la Universidad de Oviedo.
- CATTELL J. M. (1885) Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien* 2, 635-650.
- CYCOWICZ Y., FRIEDMAN D., ROTHSTEIN M., SNODGRASS J. (1997). Picture naming by young children: Norms of name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology* 65, 171-237.
- FORSTER K., FORSTER J. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers* 35, 116-124.
- GAUVIN H., DE BAENE W., BRASS M., HARTSUIKER R. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage* 126, 96-105.
- LEVELT W., ROELOFS A., MEYER A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science* 22, 1-75.
- MANOIOFF L., ARSTEIN M., CANAVOSO M., FERNANDEZ L., SEGUI J. (2010). Expanded Norms for 400 Experimental Pictures in an Argentinian Spanish-Speaking Population. *Behavior Research Methods, Instruments and Computers* 42, 452-460.
- MANOIOFF L., SEGUI J., HALLÉ P. (2013). L'effet de fréquence dans l'accès aux propriétés phonologiques des noms d'objets. *L'Année Psychologique* 113, 335-348.
- OLDFIELD R., WINGFIELD A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17, 273-281.
- ÖZDEMİR R., ROELOFS A., LEVELT W. (2007). Perceptual uniqueness point effect in monitoring internal speech. *Cognition* 105, 457-465.
- RASTLE K., DAVIS M. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance* 28, 307-3.
- RASTLE K., CROOT K., HARRINGTON J., COLTHEART M. (2005). Characterizing the motor stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance* 31, 1083-1095.
- SCHAEFFLER S., SCOBIE J., SCHAEFFLER F. (2014). Measuring reaction times: vocalization vs. articulation. *Actes de 10th ISSP*, 379-382.
- SNODGRASS J., VANDERWART M. (1980). A Standardized set of 260 pictures : Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* 6, 174-215.