

# L'étude comparative de la perception de la parole : nouveaux développements<sup>1</sup>

Franck Ramus

Laboratoire de Sciences Cognitives et Psycholinguistique  
(EHESS/CNRS/DEC-ENS)

Sous presse dans *Rééducation Orthophonique*, 2006

## Résumé

Le langage est le propre de l'homme. Cela implique-t-il que l'ensemble des capacités langagières ont évolué spécifiquement pour le langage et uniquement chez l'être humain ? Ou se pourrait-il qu'une partie de ces capacités soient plus générales et communes à d'autres espèces ? Nous présentons des travaux abordant ces questions à propos des premières étapes de la perception de la parole. Des expériences ont été menées en parallèle sur des nouveau-nés humains et sur des singes tamarins pour évaluer et comparer leur aptitude à distinguer des langues. À l'aide d'une procédure d'habituation, nous montrons que les deux populations sont capables de distinguer le néerlandais du japonais, sans exposition préalable. De plus, cette capacité est affectée lorsque la parole est jouée à l'envers, aussi bien chez le tamarin que chez le nouveau-né. Ces résultats suggèrent qu'au moins certains aspects de la perception de la parole ne sont pas nécessairement spécifiques à la parole et à l'humain, mais découlent de propriétés plus générales du système auditif des primates.

**Mots-clés:** parole, langage, prosodie, perception auditive, évolution, nouveau-nés, primates.

---

<sup>1</sup> Cet article est une version adaptée et mise à jour de Ramus, F. (2000). L'étude comparative de la perception de la parole : développements récents. *Primatologie*, 3, 421-444.

L'étude de l'apprentissage du langage par l'enfant nécessite, entre autres, de spécifier comment se forment les représentations linguistiques. Sur la base d'un nombre croissant de résultats, il est généralement admis qu'à la naissance l'enfant dispose de représentations linguistiques universelles, qui se spécialisent progressivement pour les sons de sa langue maternelle (Werker & Tees, 2005). L'étude de la perception de la parole chez le nouveau-né montre que les représentations linguistiques initiales, si elles ne sont pas spécifiques à une langue particulière, sont néanmoins remarquablement bien adaptées aux propriétés de la parole<sup>2</sup>. Leur extrême précocité rend de plus très probable qu'elles soient innées. Une question subsidiaire à celle de l'innéité est celle de la spécificité au langage: les capacités perceptives du nouveau-né sont-elles dues à l'évolution particulière de l'espèce humaine, qui les auraient sélectionnées spécialement pour analyser la parole, ou s'agit-il de capacités générales du cerveau, et notamment du système auditif, exploitées entre autres dans la perception de la parole ?

Pour répondre à cette question, Patricia Kuhl a proposé une démarche originale, consistant à tester ces mêmes capacités de perception de la parole sur des animaux :

« Lorsque les humains perçoivent la parole, nous pensons qu'ils utilisent les niveaux de traitement à la fois auditif et phonétique. Cependant, il est difficile de distinguer les deux niveaux, à la fois sur le plan théorique et sur le plan expérimental. Nous ne savons pas quels effets attribuer au niveau auditif et lesquels attribuer au niveau phonétique. Un modèle animal est utile parce qu'il résout ce problème. Si l'espèce animale est choisie convenablement (Kuhl, 1979a), elle peut fournir un bon modèle du niveau de traitement auditif de l'homme, en l'absence de tout traitement de haut niveau (phonétique). L'avantage est assez évident : l'animal reflète ce qui est naturel pour le système de traitement auditif, lorsque les influences du niveau phonétique sont éliminées et que seules restent les influences du niveau auditif » (Kuhl, 1979a, p. 360, traduction libre).

Cette démarche a engendré un vaste champ de recherches qui a permis d'éclairer la perception phonétique sous un jour nouveau. Nous allons tout d'abord en rappeler les principaux résultats, puis nous présenterons de nouvelles expériences élargissant cette approche à de nouveaux aspects de la perception de la parole.

## 1 Études comparatives de la perception phonétique

Dans les tous débuts de la recherche sur la perception de la parole, a été mis en évidence un phénomène remarquable : la perception catégorielle (Liberman, 1957). On observe typiquement ce phénomène dans le protocole suivant. Deux syllabes sont choisies, ne différant que par un trait phonétique (par exemple [da] et [ta], qui ne diffèrent que par leur VOT<sup>3</sup>), et une série de syllabes intermédiaires sont synthétisées en faisant varier régulièrement les paramètres physiques responsables de la différence (ici, le VOT). À l'écoute des syllabes intermédiaires entre [da] et [ta], les sujets ont tendance à les assimiler à l'une ou à l'autre, comme s'il existait une frontière phonétique au passage de laquelle la perception change de manière radicale et non linéaire. De plus, dans les tâches de discrimination, les sujets discriminent beaucoup moins bien les syllabes situées du même côté de la frontière que celles situées de part et d'autre, à distance acoustique égale. Ainsi, contrairement à d'autres domaines de la perception auditive, la perception phonétique n'est pas linéaire et continue,

---

<sup>2</sup> Par exemple, les nouveau-nés semblent percevoir les syllabes comme des unités pertinentes, par rapport à d'autres séquences de phonèmes (Bertoncini, Floccia, Nazzi, & Mehler, 1995; Bertoncini & Mehler, 1981; Bijeljac-Babic, Bertoncini, & Mehler, 1993).

<sup>3</sup> Voice-Onset Time : le temps entre l'explosion de la consonne et le début du voisement.

mais « catégorielle ». Cette découverte a conduit à supposer que la parole était « spéciale » (Liberman, 1982), et par conséquent nécessitait des mécanismes de traitement spécifiques. Cette supposition a été renforcée par la découverte de la perception catégorielle chez des nourrissons d'un mois, ce qui a de plus suggéré que cette capacité spécifique à la parole était innée (Eimas, Siqueland, Jusczyk, & Vigorito, 1971).

Pour tester si la perception catégorielle requiert réellement des mécanismes de traitement spécifiques à la parole, Kuhl & Miller (1975) ont testé des chinchillas<sup>4</sup>. Ils ont entraîné les chinchillas à reconnaître deux exemplaires extrêmes de [da] et [ta] (de VOT respectivement 0 et 80 ms), en utilisant une procédure de conditionnement : lorsque les animaux entendaient l'une des deux syllabes, ils devaient changer de côté dans la cage, sinon ils recevaient un léger choc électrique. Lorsqu'ils entendaient l'autre syllabe, ils devaient rester sur place, et étaient alors récompensés par un peu d'eau à boire. Après plusieurs mois d'entraînement, ils sont arrivés à des performances de discrimination supérieures à 96%. Ils ont alors été testés sur les syllabes de VOT intermédiaires entre [da] et [ta]. Il est apparu que leurs scores d'identification de ces syllabes intermédiaires étaient les mêmes que ceux de sujets humains. Ainsi, non seulement les chinchillas ont exhibé de la perception catégorielle sur le continuum [da]-[ta], mais celle-ci est organisée autour de la même frontière phonétique que les humains (un VOT d'environ 35 ms).<sup>5</sup>

D'autres expériences ont généralisé ces résultats à d'autres contrastes phonétiques, et ont pris en compte d'autres phénomènes. Notamment, la frontière phonétique correspondant au voisement n'est pas toujours au même endroit selon le lieu d'articulation : ainsi, la frontière entre [ba] et [pa] correspond à un VOT de 25 ms, alors que celle entre [ga] et [ka] correspond plutôt à 45 ms. En employant la même méthode que ci-dessus, Kuhl & Miller (1978) ont montré que les chinchillas percevaient les continuums [ba]-[pa] et [ga]-[ka] de la même manière que les humains, avec la frontière phonétique au même endroit. Ainsi, la frontière phonétique correspondant au voisement se déplace selon le contexte, de la même manière que chez l'humain.

Kuhl (1981) s'est également intéressée à la discrimination par les chinchillas de paires de syllabes sur le continuum [da]-[ta]. Elle a trouvé que comme chez l'humain, la discrimination était maximale de part et d'autre de la frontière phonétique, et beaucoup plus difficile au sein des catégories phonétiques. De tels résultats ont été répliqués chez le macaque, à la fois pour des contrastes de voisement (Kuhl & Padden, 1982) et pour des contrastes de lieu d'articulation (Kuhl & Padden, 1983).

Ces expériences ont engendré un domaine de recherches devenu trop vaste pour être traité ici exhaustivement. Notons simplement que des résultats tout aussi intéressants ont également été obtenus sur des animaux plus éloignés sur le plan phylogénétique, comme la caille (Kluender, Diehl, & Killeen, 1987) ou la perruche (Dent, Brittan-Powell, Dooling, & Pierce, 1997).

De manière générale, ces expériences ont montré qu'il n'est pas nécessaire de postuler des mécanismes qui auraient évolué spécifiquement pour le langage pour expliquer les principaux aspects de la perception phonétique. Il s'agit bien sûr de la toute première étape du traitement de la parole, et il reste probable qu'à un certain niveau de traitement plus élaboré, des mécanismes spécifiques sont bel et bien nécessaires (cf. par exemple l'effet d'« aimant perceptif », Kuhl, 1991). Le tout est de savoir à quel niveau précisément les mécanismes

---

<sup>4</sup> Le chinchilla est un rongeur vivant en Amérique du Sud.

<sup>5</sup> Voir également Morse & Snowdon (Morse & Snowdon, 1975) pour un résultat comparable sur des macaques.

spécifiques entrent en jeu. C'est dans ce but qu'a été réalisée la série d'expériences que nous présentons maintenant.

## 2 Perception prosodique par le nouveau-né humain

Si la perception phonétique est naturellement considérée comme la première étape de traitement de la parole, dans la mesure où elle concerne les unités élémentaires de parole, une autre étape de traitement tout aussi primitive et fondamentale est la perception prosodique. Le terme prosodie englobe l'ensemble des aspects supra-segmentaux de la parole, c'est-à-dire la syllabe, l'accent tonique, les tons, le rythme et l'intonation. Tout comme la phonétique, la prosodie fournit de l'information utile aux étapes ultérieures de traitement, notamment la reconnaissance des mots et l'analyse syntaxique de la phrase. Les différents modèles de l'acquisition du langage reposent de plus sur la prosodie pour initialiser l'apprentissage des mots et de la syntaxe. Cette hypothèse est compatible avec de nombreuses expériences qui ont montré une grande sensibilité des nourrissons à la prosodie.

Cette sensibilité se manifeste notamment dans des tâches de discrimination de langues. Il a en effet été montré que le nouveau-né, dès les premiers jours de vie, est capable de distinguer par exemple le français du russe, l'anglais de l'italien (Mehler et al., 1988), l'anglais de l'espagnol (Moon, Cooper, & Fifer, 1993), l'anglais du japonais (Nazzi, Bertoncini, & Mehler, 1998) ou encore le néerlandais du japonais (Ramus, Hauser, Miller, Morris, & Mehler, 2000), sans être nécessairement familier avec aucune de ces langues. Néanmoins, certaines langues sont trop proches pour être discriminées, comme l'anglais et le néerlandais (Nazzi et al., 1998). Dans quel sens sont-elles plus proches que les paires de langues précédemment citées ? Il semble qu'il s'agisse de leurs propriétés rythmiques. Les linguistes ont en effet classifié les langues en trois classes selon leur propriétés rythmiques : les langues accentuelles, parmi lesquelles le russe, l'anglais et le néerlandais, les langues syllabiques, parmi lesquelles le français, l'italien et l'espagnol, et les langues moraïques, parmi lesquelles le japonais. Ainsi, les expériences menées sur le nouveau-né peuvent être interprétées comme montrant que celui-ci discrimine les langues si elles appartiennent à des classes rythmiques différentes, mais pas si elles appartiennent à la même classe (comme l'anglais et le néerlandais)<sup>6</sup>.

Si l'hypothèse de la discrimination par le rythme est compatible avec les résultats obtenus, elle n'est cependant pas la seule possible. En effet, de nombreux indices acoustiques et linguistiques sont présents dans la parole, et d'autres indices que le rythme pourraient également rendre compte des données empiriques connues. Ce problème a été abordé par le passé en filtrant la parole (passe-bas, seuil à 400 Hz). Supprimer les fréquences élevées du signal de parole a pour effet d'éliminer l'information phonétique et par conséquent de bloquer l'accès au lexique et à la syntaxe. La prosodie, cependant est largement préservée par le filtrage. Certaines des expériences rapportées ci-dessus ont en fait été réalisées avec des phrases filtrées (Mehler et al., 1988; Nazzi et al., 1998), réduisant ainsi l'éventail des interprétations possibles : l'information prosodique est suffisante pour que le nouveau-né discrimine les langues. Récemment, nous avons essayé de déterminer le rôle respectif des deux composantes principales de la prosodie, le rythme et l'intonation. Nous avons notamment montré que lorsque des phrases néerlandaises et japonaises sont resynthétisées de manière à préserver leurs propriétés rythmiques, mais pas leur intonation<sup>7</sup>, les nouveau-nés

---

<sup>6</sup> Nous disposons également de résultats non publiés montrant que les nouveau-nés ne discriminent pas l'espagnol du catalan, ces langues étant toutes deux syllabiques.

<sup>7</sup> L'intonation des phrases était remplacée par une intonation artificielle, identique pour les deux langues.

sont encore capables de distinguer les deux langues (Ramus, 2002). L'hypothèse de la discrimination par le rythme s'en trouve donc renforcée.

Tout comme la perception catégorielle des phones, la perception de la prosodie, et notamment du rythme, fait donc probablement partie des capacités innées de base de l'être humain qui se manifestent dès la naissance. Il est dès lors légitime de se poser à propos de la perception prosodique une question identique à celle que Patricia Kuhl s'est posée à propos de la perception phonétique. S'agit-il d'une capacité qui a évolué spécifiquement pour les besoins de la perception de la parole et de l'acquisition du langage, ou s'agit-il d'une capacité plus générale du système auditif ?

Une manière d'aborder la question consiste à se demander si la perception prosodique est généralisable à des sons différents de la parole. C'est ce qu'ont fait Mehler et coll. (1988), en re-testant la discrimination français-russe sur le nouveau-né en jouant les phrases à l'envers : ils ont trouvé que la discrimination n'était plus possible dans ces conditions. La parole jouée à l'envers a bien entendu des propriétés spectrales et temporelles très proches de la parole normale, mais certains sons de la parole ont un profil énergétique asymétrique par rapport au temps. Il semble donc que la perception prosodique soit sensible à ces aspects particuliers de la parole<sup>8</sup>.

Nous avons répliqué ce résultat sur deux nouvelles langues, le néerlandais et le japonais. Les phrases ont été resynthétisées en subissant une transformation particulière appelée *saltanaj* (Ramus & Mehler, 1999) : le répertoire de phonèmes est appauvri, les fricatives étant remplacées par /s/, les voyelles par /a/, les liquides par /l/, les occlusives par /t/, les nasales par /n/ et les semi-voyelles par /j/. Cependant, les durées des phonèmes, ainsi que l'intonation de chaque phrase sont fidèlement reproduits, préservant ainsi leurs propriétés prosodiques<sup>9</sup>. Ces phrases ont été présentées à des nouveau-nés français suivant la procédure d'habituation : pendant une première phase les bébés entendent les phrases prononcées par 2 locutrices de l'une des langues puis, lorsque leur intérêt pour les stimuli commence à faiblir, ils passent à des phrases prononcées par 2 nouvelles locutrices, soit dans la même langue (groupe contrôle), soit dans l'autre langue (groupe expérimental). La réaction des bébés face aux stimuli est mesurée par leur nombre de succions par minute sur une tétine fermée, chaque succion de grande amplitude déclenchant une phrase. Consécutivement au changement, une augmentation significative des succions pour les bébés du groupe expérimental par rapport à ceux du groupe contrôle indique que les premiers ont discriminé les deux langues, indépendamment des différences entre locutrices<sup>10</sup>. Trente-deux nouveau-nés âgés de deux à cinq jours ont été exposés aux phrases jouées à l'endroit, et ont effectivement discriminé les deux langues. Trente-deux autres nouveau-nés ont été testés sur les mêmes phrases jouées à l'envers, et n'ont montré aucune réaction au changement de langue (Ramus et al., 2000).

Les raisons pour lesquelles le nouveau-né humain ne traite pas la parole à l'envers de la même manière que la parole à l'endroit ne sont pas encore élucidées. Plusieurs hypothèses sont

---

<sup>8</sup> D'autres expériences ont par ailleurs montré que la parole à l'envers n'active pas les mêmes zones du cerveau que la parole à l'endroit, même si cette dernière est dans une langue inconnue (Neville & Mills, 1997; Perani et al., 1996).

<sup>9</sup> Des exemples de stimuli sont disponibles à l'adresse Internet suivante : <http://www.lscp.net/persons/ramus/resynth/ecoute.htm>

<sup>10</sup> L'utilisation de plusieurs locutrices par langue et la comparaison avec le groupe contrôle permet donc de montrer que la discrimination de langues ne se réduit pas à une discrimination de voix.

envisageables. Une première hypothèse est que l'être humain possède, dès la naissance, un filtre dans les entrées auditives qui aiguille les sons de parole vers les étapes supérieures du traitement linguistique, ce qui éviterait d'inonder les processus linguistiques d'une multitude de bruits non pertinents. Ce filtre devrait opérer sur la base de propriétés spectrales et/ou temporelles caractéristiques de la parole. Une autre hypothèse fait l'économie d'un filtre, mais suppose que c'est l'extraction du rythme de la parole qui repose sur certaines propriétés spécifiques. Par exemple, un modèle de l'extraction du rythme repose sur la segmentation de la parole en consonnes et voyelles (Ramus, Nespors, & Mehler, 1999) ; cette segmentation pourrait être altérée si certaines consonnes ne sont pas reconnues comme telles lorsque la parole est jouée à l'envers. Un autre modèle de la perception du rythme repose explicitement sur les profils d'énergie acoustique des syllabes (Howell, 1988; Morton, Marcus, & Frankish, 1976), et ceux-ci sont susceptibles d'être altérés lorsque la parole est jouée à l'envers. Le point commun de ces deux hypothèses est qu'elles supposent que le nouveau-né exploite des propriétés du signal spécifiques à la parole humaine, et que ces propriétés sont altérées lorsque la parole est jouée à l'envers. Le traitement différentiel par le nouveau-né de la parole normale et de la parole à l'envers constitue donc un argument en faveur de la spécificité, suggérant que l'appareil perceptif du nourrisson est finement adapté aux propriétés de la parole, du moins en ce qui concerne la perception prosodique.

Les autres espèces animales n'ont, bien sûr, aucune raison d'avoir des capacités perceptives spécifiques aux propriétés de la parole humaine. Elles ont, en revanche, des capacités auditives qui peuvent être remarquables, et qui sont éventuellement adaptées à leurs propres vocalisations (Ghazanfar & Hauser, 1999). En l'absence de données sur la perception prosodique par d'autres espèces, les prédictions restent ouvertes. On pourrait prédire que d'autres espèces seraient incapables de discriminer le néerlandais du japonais, mais on pourrait tout aussi bien prédire le contraire : certaines espèces pourraient tout à fait être sensibles aux aspects rythmiques et/ou mélodiques des signaux sonores. Dans ce cas, on pourrait de plus prédire que, pour des organismes qui ne sont pas spécifiquement adaptés à la parole, la discrimination devrait être aussi facile à l'envers qu'à l'endroit. Les expériences qui suivent visent à tester ces prédictions.

### **3 Discrimination de langues par des tamarins**

Nos expériences sur les singes ont été planifiées simultanément avec les expériences correspondantes chez le nouveau-né, l'idée étant de comparer directement le comportement des nouveau-nés et des singes dans des expériences similaires et sur des stimuli identiques. Les expériences décrites ci-dessous ont été réalisées avec la collaboration de Marc Hauser, Cory Miller et Dylan Morris, au Primate Cognitive Neuroscience Laboratory, Harvard University (Ramus et al., 2000).

#### **3.1 Discrimination néerlandais-japonais, à l'endroit et à l'envers**

Avant d'utiliser la parole resynthétisée, nous avons préféré faire une première série d'expériences avec de la parole naturelle, préservant ainsi le maximum d'indices pour la discrimination. Nous avons donc utilisé les mêmes phrases que celles utilisées avec les nouveau-nés, dans leur version originale, puis jouées à l'envers. Les sujets étaient 13 tamarins << cotton-top >> (*saguinus oedipus oedipus*), adultes, nés en captivité, et résidant au Primate Cognitive Neuroscience Laboratory.

##### **3.1.1 Stimuli**

Les phrases ont été extraites du corpus multilingue du LSCP (Nazzi, 1997; Nazzi et al., 1998). Elles ont été lues par quatre locutrices natives du néerlandais, et quatre du japonais. Cinq phrases par locutrices ont été choisies, constituant un total de vingt phrases par langues. Les phrases choisies étaient appariées en nombre de syllabes et en durée.

### 3.1.2 Méthode

Nous avons utilisé une procédure d'habituation/déshabituaiton similaire dans son principe à celle utilisée sur les nouveau-nés, mais en prenant comme mesure expérimentale l'orientation du regard plutôt que la succion.

Huit conditions ont été obtenues par le croisement des facteurs groupe (contrôle/expérimental), stimuli (endroit/envers) et langue d'habituation (néerlandais/japonais). Dû au faible nombre de sujets, chacun a participé à 4 conditions différentes (groupe  $\times$  stimuli), et ils ont été répartis en deux groupes pour contrebalancer la langue d'habituation. L'ordre de passage des conditions a été contrebalancé à travers les sujets.

Le test se déroule dans une cabine isolée du reste de la colonie. Le tamarin est installé dans une cage au-dessus de laquelle, dans un coin, se trouve un haut-parleur dissimulé. Le test est filmé par une caméra vidéo qui englobe toute la cage dans son champ. Deux expérimentateurs suivent le déroulement du test sur un moniteur situé à l'extérieur de la cabine.

Le principe est qu'à la diffusion d'une phrase par le haut-parleur, le tamarin se retourne et regarde dans cette direction. Puis, rien de nouveau ne se passant, il reprend son activité (qui consiste à explorer la cage). À force de diffuser des phrases, le tamarin s'habitue, et ne se retourne plus vers le haut-parleur. Si on diffuse alors une phrase dans une nouvelle langue, va-t-il se déshabituer et se retourner à nouveau ? Si c'est le cas, on aura une bonne indication que le tamarin a décelé une différence entre cette langue et celle à laquelle il s'était habitué.

Un essai est lancé par l'expérimentateur à un moment où le tamarin regarde vers le bas et n'est pas face au haut-parleur. L'essai consiste en la diffusion d'une phrase. Une réponse positive de la part du singe consiste à s'orienter vers le haut-parleur avant la fin de la diffusion de la phrase (environ 3 secondes). Le début et la fin de l'orientation vers le haut-parleur sont codés en temps réel par l'expérimentateur. Si le tamarin ne s'oriente pas vers le haut-parleur, ou s'oriente après la fin de la phrase, la réponse est codée comme négative. Lorsqu'une orientation est ambiguë, il n'est pas tenu compte de l'essai pour le critère d'habituation.

Pendant la phase d'habituation, les phrases correspondant à la condition sont diffusées dans un ordre aléatoire, comme pour les bébés. Le critère d'habituation est atteint lorsque le singe ne s'oriente pas vers le haut-parleur lors de deux essais consécutifs. Il est alors « habitué » aux stimuli. Un minimum de 2 phrases d'habituation doivent également avoir été entendues (et suivies d'une orientation vers le haut-parleur) pour que le critère soit atteint.

La phase de test consiste en deux essais, diffusant chacun une phrase de test. Seul le premier essai de test est pris en compte pour déterminer si le singe s'est déshabitué ou pas. Dans l'éventualité où le singe ne réagit pas au cours du premier essai de test, un essai de post-test est lancé : il consiste en la diffusion d'un cri long de tamarin<sup>11</sup>. Le post-test sert à vérifier si le tamarin ne s'est pas habitué, non seulement aux stimuli d'habituation, mais à toute la session

---

<sup>11</sup> Il s'agit d'un cri servant à établir le contact entre les individus.

expérimentale. S'il ne réagit pas au post-test, on considère donc qu'il n'était pas en état de réagir aux stimuli de test, et les résultats de cette session sont rejetés. Lorsque l'orientation vers le haut-parleur est ambiguë pour l'une des deux phrases de test ou pour le post-test, les résultats sont également rejetés. Après la session, les essais de test sont recodés par deux observateurs d'après la bande vidéo, sans connaître la condition expérimentale.

### 3.1.3 Résultats

La Figure 1 donne le nombre de sujets ayant répondu positivement et négativement au premier essai de test, en fonction du type de stimuli (à l'endroit ou à l'envers), et du type de changement (de langue ou de locutrices).

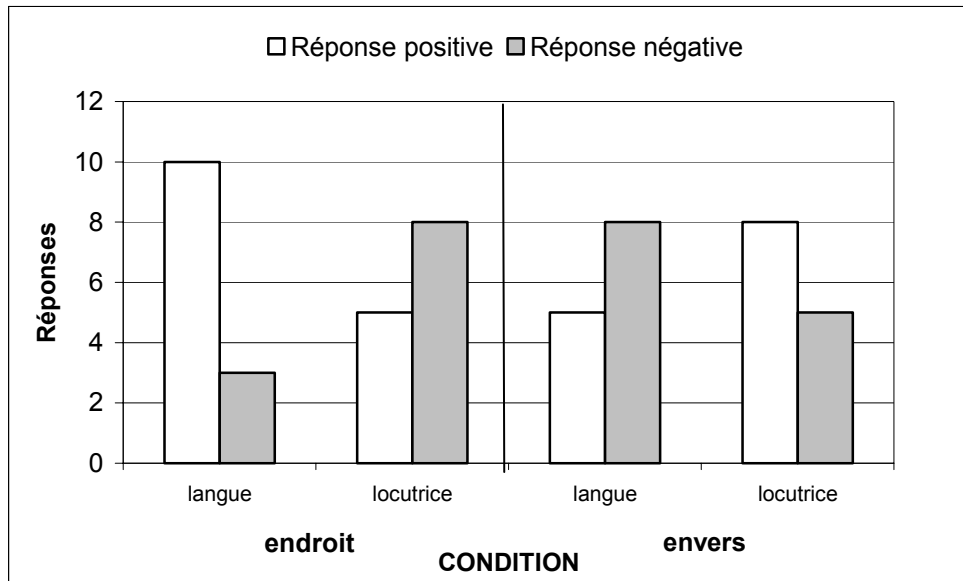


Figure 1: Discrimination néerlandais-japonais par les tamarins, en parole naturelle à l'endroit et à l'envers. Adapté de Ramus et al. (2000).

Dans la condition changement de langue à l'endroit, 10 tamarins sur 13 se sont déshabitués, ce qui est significativement différent du hasard ( $p < 0.05$  par un test binomial). Mais comme pour les nouveau-nés, cela ne signifie rien si l'on ne prend pas en compte le groupe contrôle, qui ne changeait que de locutrices. Dans ce groupe, une majorité de tamarins ne s'est pas déshabituée. La différence entre les deux groupes est significative ( $\chi^2 = 3.94$ ,  $p < 0.05$ ), montrant que les tamarins se sont plus déshabitués au changement de langues qu'au changement de locutrices. On peut donc en conclure que les tamarins ont discriminé le néerlandais du japonais.

En revanche, dans les conditions où les phrases étaient jouées à l'envers, la situation est tout à fait différente : une majorité de tamarins ne s'est pas déshabituée au changement de langue, mais s'est déshabituée au changement de voix, sans que la tendance soit significative. De plus il n'y a pas de différence entre le changement de langue et le changement de voix ( $\chi^2 = 1.38$ ,  $p > 0.25$ ). Les tamarins n'ont donc pas discriminé les deux langues lorsque les phrases étaient jouées à l'envers.

### 3.1.4 Discussion



L'expérience ci-dessus nous a apporté deux résultats remarquables. Le premier, c'est que des animaux ayant un système auditif comparable au nôtre, les tamarins, peuvent discriminer deux langues comme le néerlandais et le japonais. Ceci suggère que l'aptitude du nouveau-né à distinguer les langues pourrait reposer sur des capacités auditives générales, non spécifiques au langage.

Le deuxième résultat remarquable, c'est que cette capacité des tamarins à discriminer les langues ne se maintient pas lorsque les phrases sont jouées à l'envers. Ceci suggère que nous devons expliquer le même comportement chez le nouveau-né sans postuler un traitement spécifique à la parole.

Bien entendu, on aimerait pouvoir préciser *comment* le tamarin est parvenu à distinguer le japonais du néerlandais. Peut-être y parvient-il par des moyens très différents de ceux du nouveau-né, auquel cas il y aurait peu de conclusions à en tirer en ce qui concerne la capacité du nouveau-né. Une expérience supplémentaire a été conduite avec de la parole resynthétisée comme pour le nouveau-né, mais n'a pas abouti à des résultats concluants. Nous avons donc été conduits à revenir à la parole naturelle et à élargir plutôt l'éventail des langues considérées.

### **3.2 Discrimination polonais-japonais et anglais-néerlandais, à l'endroit et à l'envers**

Compte tenu de petit nombre de sujets testés dans les expériences ci-dessus, il était légitime de s'interroger sur la fiabilité des résultats, et donc de tenter de les répliquer. C'est ce qui a été fait lors d'une nouvelle étude sur 17 tamarins (Tincoff et al., 2005). Cette fois deux nouvelles paires de langues ont été testées : deux langues de classes rythmiques différentes (Polonais/Japonais), et deux langues de même classe rythmique (Anglais/Néerlandais). Lors de différentes sessions, elles ont été jouées à nouveau soit à l'endroit, soit à l'envers. Les phrases étaient sous leur forme naturelle, prononcées par 4 locutrices par langue. Les résultats sont présentés Figure 2. On peut voir que les tamarins ont distingué le polonais du japonais, lorsque ces langues étaient jouées à l'endroit, mais pas à l'envers. Ceci confirme les résultats précédemment obtenus, avec une nouvelle langue, le polonais. De plus, les tamarins n'ont pas distingué l'anglais du néerlandais, bien qu'ils soient très familiers avec l'anglais (ils sont exposés quotidiennement).

Ainsi, l'ensemble de ces données suggère que les tamarins sont bien capables de distinguer des langues humaines, au moins lorsque celles-ci appartiennent à des classes rythmiques différentes. De plus, leur absence de réaction aux différences entre anglais et néerlandais (qui possèdent des différences phonétiques, mais pas rythmiques), renforce l'hypothèse selon laquelle les tamarins, comme les nouveau-nés humains, seraient sensibles au rythme de la parole.

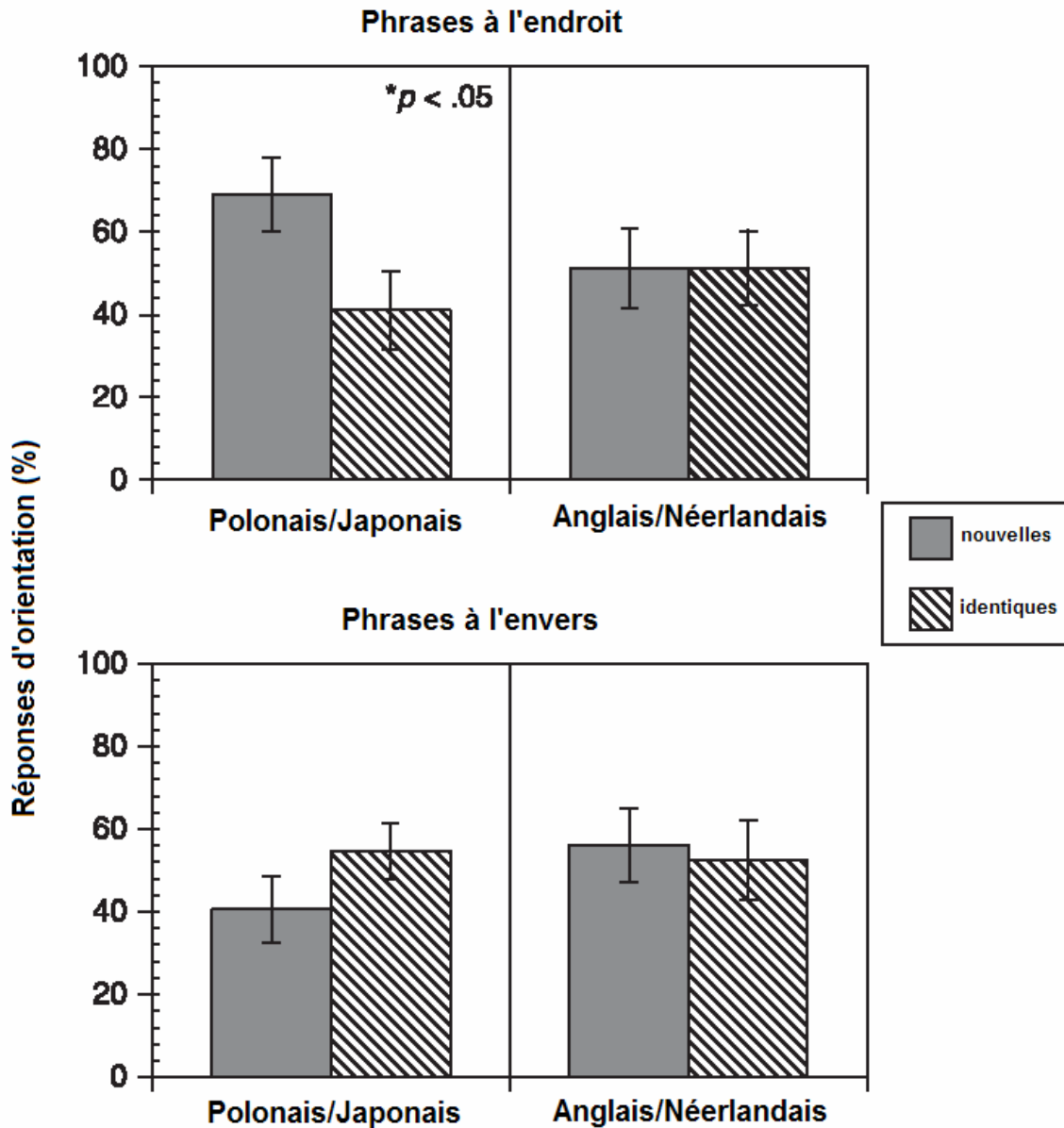


Figure 2: Discrimination polonaise-japonaise et anglaise-néerlandaise par les tamarins, à l'endroit ou à l'envers. Adapté de Tincoff et coll. (2005).

## 4 Discussion générale

Ces expériences apportent selon nous un certain nombre d'avancées significatives par rapport aux précédentes études comparatives de la perception de la parole:

- D'un point de vue méthodologique, l'utilisation de la procédure d'habituation/déshabituement présente plusieurs avantages. Rappelons en effet que la procédure de conditionnement utilisée par Kuhl demandait aux chinchillas et aux macaques plusieurs mois d'entraînement à la tâche sur les exemplaires extrêmes du continuum phonétique. Sans parler du coût et de la lourdeur de telles expériences, on peut se demander à quel point cette exposition permanente et prolongée aux phonèmes n'a pas sensibilisé le système auditif de ces animaux. La procédure

d'habituation/déshabituation, elle, ne requiert aucune exposition préalable, et teste directement les réactions spontanées des singes face aux stimuli. Les conclusions que l'on peut en tirer ont donc d'autant plus de portée. Le fait que cette procédure soit la même (sur le principe) que celle utilisée sur les nouveau-nés facilite également la comparaison directe entre les deux populations. En revanche, elle limite le nombre de mesures que l'on peut obtenir par sujets, ce qui explique sans doute qu'elle ait été moins utilisée par le passé (mais voir Morse & Snowdon, 1975).

- Les études comparatives de la perception de la parole se sont jusqu'à présent limitées à la perception de sons et de syllabes isolées. En étudiant la perception de phrases entières, nos expériences s'adressent à une classe plus étendue et plus générale de processus perceptifs.
- Enfin, Kuhl (1987) a suggéré que l'une des caractéristiques du traitement de la parole par l'humain qui était susceptible de ne pas être retrouvée chez le singe était la « perception de classes d'équivalence auditives ». Elle faisait référence à la capacité du nourrisson (et de l'adulte) à considérer comme équivalentes des formes acoustiques différentes du même phonème, à savoir le même phonème prononcé par plusieurs locuteurs (Kuhl, 1979b, 1983). Ici, nous avons montré que les singes discriminaient le néerlandais du japonais à travers les voix de plusieurs locutrices, et réagissaient plus au changement de langue qu'au changement de voix. Il s'agit donc d'une toute première indication de la perception de classes d'équivalence auditives par le singe.

Partant de l'observation que l'être humain possède, dès la naissance, une sensibilité innée pour le rythme de la parole, nous nous sommes posé la question de savoir si cette sensibilité était une fonction cognitive ayant évolué spécifiquement pour le langage, ou si elle découlait des propriétés générales de notre système auditif. L'incapacité du nouveau-né à discriminer les langues lorsque les phrases sont jouées à l'envers semblait être un bon argument en faveur de la spécificité. Ayant découvert qu'il en est de même chez le tamarin, nous sommes amenés à reconsidérer cet argument.

Les deux types d'hypothèses que nous avons retenues à la Section 2 pour expliquer le comportement du nouveau-né faisaient appel à des processus spécifiques à la parole. Il est néanmoins possible de re-formuler ces hypothèses d'une manière plus générale et plus pertinente dans le cas du tamarin. On peut en effet imaginer que tout organisme doué d'un système auditif doit faire le tri entre les sons qui sont pertinents pour lui et tous les bruits qui ne le sont pas ; ou plus généralement, que les procédures de traitement auditif doivent être particulièrement adaptées aux propriétés spécifiques de ces « sons pertinents » pour l'organisme. Ces sons étant principalement ceux émis par ses congénères, ses proies et ses prédateurs, il est probable que les sons pertinents pour un grand nombre d'espèces aient en commun un certain nombre de propriétés acoustiques caractéristiques (périodicité, harmonicité, répartition de l'énergie spectrale en formants...). On pourrait ainsi postuler que tous les primates (et peut-être tous les mammifères) ont une capacité innée à reconnaître les sons émis par les autres animaux (ou au moins leurs vocalisations), et une capacité de traitement accrue pour ceux-ci (Seyfarth & Cheney, 1997). La parole à l'envers pourrait ne pas respecter certaines propriétés typiques de ces sons, et être ainsi rejetée par le système auditif d'animaux non humains. Cette hypothèse, qui mériterait d'être testée plus en détail, et sur un large éventail d'espèces, ouvre la porte à des considérations nouvelles sur l'évolution du langage.

L'apparition du langage au sein de l'espèce humaine est un sujet hautement épineux et controversé. La raison principale en est certainement l'absence quasi-totale de données

empiriques. Contrairement aux organes biologiques, les capacités cognitives ne laissent pas de traces dans les fossiles<sup>12</sup>; et contrairement à d'autres capacités cognitives, le langage humain n'a pas d'équivalent ni de proche précurseur parmi les autres espèces<sup>13</sup>. Aussi peu contraints par les données, les spéculations et scénarios hypothétiques de l'évolution du langage vont bon train. Un certain nombre de ces scénarios se concentrent sur un sous-problème de l'évolution du langage, celui de l'émergence de la parole<sup>14</sup> (MacNeilage, 1998; Rizzolatti & Arbib, 1998). Il s'agit de comprendre comment le tract vocal de nos ancêtres a évolué de manière à produire la grande variété de sons qui caractérise la parole humaine. Néanmoins, il est tout aussi important de comprendre comment notre appareil perceptif peut sembler aussi finement spécialisé pour traiter et catégoriser les sons de parole. Dans les scénarios sus-cités, cet aspect est souvent implicitement considéré comme secondaire à l'émergence des capacités de production. Nos travaux, comme ceux de Kuhl, tendent à montrer que certaines de ces capacités perceptives existaient préalablement à l'évolution du tract vocal humain. Ceci suggère que, au moins pour un certain nombre d'aspects de la parole, c'est l'appareil de production qui a dû s'adapter aux domaines de sensibilité de notre appareil perceptif, plutôt que le contraire (voir également Kluender, 1994; Kuhl, 1988; Stevens, 1989). De nombreux aspects de la perception de la parole n'ayant pas encore été explorés de manière comparative, il va de soi que cette conclusion n'est pas immédiatement généralisable.

Jusqu'où peut-on aller dans l'homologie entre perception de la parole et perception auditive chez les primates non humains ? De ce point de vue, les études purement comportementales sont intrinsèquement limitées: la similarité de comportements entre deux espèces n'implique pas nécessairement que ces comportements sont engendrés par des fonctions cognitives identiques. Nous avons déjà mentionné qu'il n'est pas certain que les tamarins discriminent les langues sur la base des mêmes indices que les nouveau-nés humains. En ce qui concerne les difficultés similaires qu'ont les deux populations avec la parole et l'envers, on pourrait également imaginer que les tamarins rejettent celle-ci pour d'autres raisons que les nouveau-nés. Cette hypothèse est parfaitement testable : il est en effet possible de resynthétiser les phrases néerlandaises et japonaises en isolant ou en supprimant les propriétés acoustiques de la parole à l'envers dont on pense qu'elles posent problème, et d'observer dans quelles conditions la discrimination est ou non possible pour chacune des deux espèces. Dans l'hypothèse où les deux espèces manifestent le même comportement dans les différentes conditions, l'homologie sera probable. Même dans ce cas, des études neurophysiologiques seraient bienvenues pour confirmer que des structures neuronales similaires sont à la base des comportements homologues.

C'est en fait l'objet d'une nouvelle collaboration internationale récemment financée par la Commission Européenne, le projet Neurocom (<http://neurocomm.free.fr/>). Il s'agit notamment de conduire un certain nombre d'études d'imageries cérébrales à la fois sur des adultes, des nourrissons et des singes (des macaques). Au cours de ces expériences, les sujets sont exposés à du langage humain (des phrases de leur langue maternelle ou d'une langue étrangère), à des cris humains émotionnels sans langage, à des vocalisations de macaques, à des chants d'oiseaux et à des stimuli contrôlés. La principale question qui est posée dans ce projet est celle des origines évolutives des aires cérébrales spécialisées pour le langage chez l'humain. Une hypothèse est qu'elles sont construites principalement sur la base des aires spécialisées

---

<sup>12</sup> Mis à part la taille et la configuration de la boîte crânienne.

<sup>13</sup> Par exemple, l'assemblage productif de mots en phrases selon des règles syntaxiques abstraites est absent de tous les autres systèmes de communication animale.

<sup>14</sup> Il s'agit d'un sous-problème dans la mesure où une explication de l'évolution de la parole ne suffirait pas à expliquer celle par exemple de la syntaxe.

pour la perception des vocalisations chez le singe. Evidemment, au niveau des aires auditives les plus primaires des recouvrements sont attendus. Mais au-delà, une hypothèse alternative serait que les vocalisations des singes ne constitueraient pas un véritable précurseur du langage, mais seraient plutôt les ancêtres des vocalisations émotionnelles (non linguistiques) humaines. Une autre hypothèse est celle selon laquelle le langage humain tirerait ses principales propriétés du système moteur plutôt qu'auditif, et notamment de « neurones miroirs » spécialisés dans l'imitation vocale (Rizzolatti & Arbib, 1998). Une exploration électrophysiologique des neurones miroirs potentiellement reliés à la perception et à la production des vocalisations chez le macaque est justement incluse dans le projet Neurocom. Enfin, il reste bien sûr la possibilité selon laquelle les principales caractéristiques du système cérébral pour le langage ne sont directement héritées ni du système auditif, ni du système moteur des primates, et résulteraient de spécialisations cérébrales encore distinctes.

Quelle que soit l'hypothèse, il reste clair qu'il existe des niveaux de traitement linguistique spécifiquement humains, que l'on ne retrouvera pas chez le singe. Même s'il apparaît que la totalité des aspects universels de la perception de la parole sont partagés avec les singes, il n'en restera pas moins que l'usage que font les deux espèces de ce traitement auditif sera fondamentalement différent. Les tamarins et les nouveau-nés humains peuvent bien avoir les mêmes capacités perceptives, seuls ces derniers s'en serviront pour apprendre la phonologie, la syntaxe et le lexique de leur langue maternelle.

## Références

- Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech*, 38, 311-329.
- Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant perception. *Infant Behavior and Development*, 4, 247-260.
- Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, 29, 711-721.
- Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, 102(3), 1891-1897.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Ghazanfar, A. A., & Hauser, M. D. (1999). The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends Cogn Sci*, 3(10), 377-384.
- Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope: II. *Perception & Psychophysics*, 43, 99.
- Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173-217). San Diego: Academic Press.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195-1197.
- Kuhl, P. K. (1979a). Models and mechanisms in speech perception: Species comparison provide further contributions. *Brain, behavior and evolution*, 16, 374-408.
- Kuhl, P. K. (1979b). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66, 1668-1679.

- Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, 70, 340-349.
- Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6, 263-285.
- Kuhl, P. K. (1988). Auditory perception and the evolution of speech. *Human Evolution*, 3, 19-43.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept Psychophys*, 50(2), 93-107.
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
- Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63, 905-917.
- Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, 32, 542-550.
- Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, 73, 1003-1010.
- Lieberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117-123.
- Lieberman, A. M. (1982). On finding that speech is special. *American Psychologist*, 37, 148-167.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav Brain Sci*, 21(4), 499-511; discussion 511-546.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495-500.
- Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics*, 17, 9-16.
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83(5), 405-408.
- Nazzi, T. (1997). *Du rythme dans l'acquisition et le traitement de la parole*. Unpublished doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756-766.
- Neville, H. J., & Mills, D. L. (1997). Epigenesis of language. *Mental retardation and developmental disabilities research reviews*, 3, 282-292.
- Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *NeuroReport*, 7, 2439-2444.
- Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In B. Bel & I. Marlien (Eds.), *1st International Conference on Speech Prosody* (pp. 115-120). Aix-en-Provence.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 349-351.

- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105(1), 512-521.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci*, 21(5), 188-194.
- Seyfarth, R. M., & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 249-273). Cambridge: Cambridge University Press.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
- Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: Further tests with a nonhuman primate. *Developmental Science*, 8(1), 26-35.
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev Psychobiol*, 46(3), 233-251.