



Comment étudier l'influence des stéréotypes de genre et de race sur la perception de la parole ?

Aron Arnold, Maria Candea

► To cite this version:

Aron Arnold, Maria Candea. Comment étudier l'influence des stéréotypes de genre et de race sur la perception de la parole ?. *Langage et Société*, 2015, pp.75-96. halshs-01158437

HAL Id: halshs-01158437

<https://shs.hal.science/halshs-01158437>

Submitted on 1 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Version auteur

Comment étudier l'influence des stéréotypes de genre et de race sur la perception de la parole ?

Arnold, Aron, UMR 7018, CNRS & Université Sorbonne nouvelle
aron.arnold@univ-paris3.fr

Candea, Maria, CLESTHIA, Université Sorbonne nouvelle
maria.candea@univ-paris3.fr

note de la 1^{re} page

*Nous remercions Eric Bilinski, du laboratoire LIMSI-CNRS pour les parties programmation et interface graphique de cette expérimentation. Cette collaboration a été rendue possible grâce à l'accueil de M. Candea au LIMSI entre 2011 et 2013.

Résumé

Cette étude interroge les méthodes expérimentales traditionnellement utilisées en phonétique pour étudier la perception des attitudes/émotions par la voix. Nous avons voulu étudier l'effet des stéréotypes de *genre* et de *race* sur la perception, à travers un protocole basé sur les techniques d'*amorçage* et du *locuteur masqué*. Nous avons ainsi testé l'approbation de phrases évaluatives appariées à de courts extraits sonores par deux groupes indépendants de 55 personnes chacun. Chaque phrase contenait un prénom féminin ou masculin, français ou maghrébin pour amorcer les catégories identitaires « femme », « homme », « Français de souche », « d'origine arabe ». Les résultats obtenus ne sont pas toujours conformes à ceux des études antérieures menées surtout aux Etats-Unis et laissent penser que les réactions des auditrices/eurs face aux stéréotypes de genre sont moins réfléchies que les réactions face aux stéréotypes raciaux. Les conclusions incitent à mieux prendre en compte l'interaction entre chercheurs et enquêtés lors d'expériences perceptives, et à interpréter avec plus de prudence les résultats produits par des méthodes expérimentales purement quantitatives.

Mots clés : sociophonétique, perception, méthodes quantitatives, méthodes mixtes, stéréotypes de genre, stéréotypes de race, études françaises.

Abstract

This study questions the validity of experimental methods traditionally used in phonetics. We wanted to investigate the influence of gender and race stereotypes on the perception of attitudes and emotions through voice. Thus we used a method combining *priming* and *matched guise* technique to test how two independent groups of 55 subjects responded to evaluative sentences that were presented while they listened to short sound excerpts. Each sentence contained a feminine or masculine, french or maghrebian first

name in order to prime the categories "woman", "men", "French", "Arabic". Interestingly, our results don't match those of previous studies and lead us to think that the reactions of the subjects towards gender stereotypes are less controlled than the reactions towards race stereotypes. We conclude that (1) closer attention must be paid to the interaction between researcher and subjects during experimental studies and that (2) cautiousness should be exercised in the interpretation of results produced through solely quantitative research methods.

keywords: sociophonetics, perception, quantitative methods, mixed methods, gender stereotypes, race stereotypes, French studies

Les données que nous analysons dans ce texte sont issues d'une étude exploratoire dont l'objectif est d'approfondir les connaissances sur la manière dont les stéréotypes de *genre*¹ et de *race*² influencent la perception de la parole.

Alors que, par exemple, aux États-Unis ce sujet a fait l'objet d'un certain nombre de travaux en sociolinguistique, sociologie ou anthropologie (par exemple Niedzielski, 1999 ; Rubin, 1992 ; Strand & Johnson, 1996), peu d'études l'ont abordé en phonétique, en France. Une des raisons principales de son absence est sans doute la persistance du mythe d'une expérimentation objective et axiologiquement neutre qui va de pair avec la négation ou l'ignorance de l'influence des représentations stéréotypiques sur la perception. Par exemple, l'influence des stéréotypes de genre ou de race que les auditeurs et auditrices projettent sur les voix données à écouter lors d'expériences perceptives n'est que rarement, voire jamais, thématisée.

Nous avons souhaité mettre à l'épreuve la méthode expérimentale habituellement utilisée en phonétique en nous inspirant des travaux cités *supra*. Pour ce faire, nous avons mis en place un protocole de test de perception qui mobilise les deux catégories, genre et race, de manière imbriquée. Cependant, notre étude a abouti à un certain nombre de résultats qui ne coïncident pas exactement avec les résultats des travaux antérieurs, ce qui nous conduit à interroger la divergence entre l'influence des stéréotypes de race et ceux de genre ainsi que ses conséquences sur la validité des approches expérimentales exclusivement quantitatives.

¹ Le genre peut être défini comme « le dispositif par lequel le masculin et le féminin sont produits et normalisés en même temps que les formes interstitielles hormonales, chromosomiques, psychiques et performatives du genre » (Butler, 2006 : 59). Pour une discussion critique du concept de genre, voir par exemple Butler (1990, 1993, 2004) ou Dorlin (2008).

² Nous utilisons le concept de *race* au sens de construction sociale mobilisant la fiction d'un fondement biologique. La racialisation d'un individu ou d'un groupe, la construction de son l'altérité, peut se fonder sur des différences de traits somato-morphologiques, réelles ou fantasmées, comme elle peut se fonder sur des différences culturelles, socioéconomiques, d'origines régionales ou ethniques. Dans le cas précis de notre étude, le concept de *race* permet de garder au premier plan le lien avec le *racisme*, lien qui serait fortement escamoté si nous utilisions le concept plus consensuel d'*ethnicité*. Pour une discussion critique, voir par exemple Dorlin (2006), Guillaumin (1992), Lopez (1994) ou Monneyron et Siary (2012).

1. Stéréotypes et perception

Les études sur la cognition menées au cours des dernières décennies ont permis de montrer que la perception humaine ne repose pas uniquement sur un traitement ascendant (*bottom-up*), mais aussi sur un traitement descendant (*top-down*) : la perception n'est pas simplement une opération de recueil d'informations par les organes sensoriels, mais un processus dans lequel les informations recueillies sont structurées par les représentations stockées en mémoire (Bruner, 1957 ; Gregory, 1980 ; Rock, 1983). Si nos représentations exercent une telle influence sur notre perception, alors les stéréotypes – ces « croyances sur les caractéristiques, attributs et comportements des membres d'un certain groupe » (Hilton & Von Hippel, 1996 ; notre traduction) – doivent jouer un rôle déterminant dans notre manière de percevoir le monde qui nous entoure (Hippel et *al.*, 1995 : 181).

Des études en psychologie sociale ont confirmé que les représentations stéréotypiques influencent la perception des comportements des individus. Geis (1993 : 12) explique que l'observateur se focalise sur certains comportements attendus et porte beaucoup moins d'attention aux comportements non-stéréotypiques. Elle écrit que les stéréotypes « amplifient les perceptions, interprétations et mémoires qui coïncident avec les attributs stéréotypiques et obscurcissent, dissipent ou provoquent une non-prise en compte ou un oubli des informations qui ne coïncident pas avec ceux-ci » [notre traduction]. De cette manière, des comportements identiques peuvent être perçus différemment en fonction des représentations stéréotypiques de genre ou de race que l'observateur va projeter sur l'observé. Un exemple devenu classique est celui de l'étude de Condry et Condry (1976). Ces derniers ont remarqué qu'au cours d'une expérience perceptive, un nourrisson qui pleurait était perçu comme étant en colère quand les observateurs pensaient qu'il s'agissait d'un garçon et comme étant anxieux quand les observateurs pensaient qu'il s'agissait d'une fille. Dans cette expérience, les observateurs percevaient donc des états émotionnels différents alors qu'ils observaient tous exactement le même comportement, le même nourrisson et la même scène.

Cette influence des représentations stéréotypiques a aussi été attestée dans la perception de la voix et de la parole. Différentes études phonétiques ont permis de montrer que des sons pouvaient être perçus différemment en fonction du genre assigné à la personne écoutée (Johnson et *al.*, 1999 ; Strand & Johnson, 1996 ; Strand, 1999, 2000). Par exemple, Strand et Johnson (1996) ont réalisé une expérience dans laquelle ils faisaient écouter des sons – des fricatives sourdes – associées à des photos de visages féminins ou masculins pour produire l'impression que les sons entendus étaient produits tantôt par une femme, tantôt par un homme. Ils ont remarqué qu'un même son pouvait être perçu comme une fricative sourde post-alvéolaire /ʃ/ s'ils présentaient la photo d'un visage féminin et comme une fricative sourde alvéolaire /s/ s'ils présentaient un visage masculin. Les auteurs expliquent ce phénomène par des attentes spécifiques des auditrices et auditeurs par rapport aux frontières entre /s/ et /ʃ/ en fonction du genre du locuteur – on s'attend à ce que le bruit de friction soit réalisé dans des fréquences plus élevées chez une femme que chez un homme.

D'autres études ont montré qu'il y avait aussi une influence des stéréotypes de

race sur la perception de la parole. Par exemple, Niedzielski (1999) a mené une expérience perceptive portant sur des diphtongues auprès de personnes originaires de Detroit (Michigan). À la moitié des auditeurs, Niedzielski faisait écouter un enregistrement de parole en prétendant qu'il s'agissait d'une locutrice étasunienne du Michigan. À l'autre moitié, elle faisait écouter le même enregistrement en prétendant qu'il s'agissait d'une locutrice canadienne de l'Ontario. Les auditeurs devaient ensuite choisir parmi un ensemble de diphtongues créées par synthèse vocale celles qui se rapprochaient le plus des diphtongues produites par la locutrice. Avec cette expérience, Niedzielski a remarqué qu'en fonction de l'origine présumée de la locutrice, les auditeurs avaient des perceptions différentes des diphtongues présentées. Par exemple, s'ils pensaient que la locutrice était canadienne, ils tendaient à choisir les diphtongues plus ouvertes, typiques de l'anglais canadien. Cette différence de perception était donc provoquée par la représentation que les auditeurs américains avaient des Canadiens – comme appartenant à une communauté linguistique différente.

L'étude menée par Rubin (1992) constitue un autre exemple de l'influence des stéréotypes raciaux sur la perception. Dans cette étude, elle a fait écouter à deux groupes d'étudiants et étudiantes les mêmes enregistrements de la parole d'une locutrice originaire de l'Ohio, ayant l'anglais comme langue première. Ces enregistrements ont été présentés comme des extraits de cours donnés par une enseignante. Pendant l'écoute, une photo d'une femme « asiatique » a été montrée à un groupe et une photo d'une femme « caucasienne » à l'autre. Des questionnaires et tests de compréhension sous forme de texte à trous ont permis à Rubin de remarquer que lorsque l'enregistrement était présenté comme produit par une enseignante « asiatique », les étudiants percevaient un fort accent étranger et éprouvaient des difficultés à comprendre ce qui était dit.

En nous inspirant de ces différentes études réalisées aux États-Unis, nous avons souhaité tester expérimentalement s'il était possible d'attester cette même influence des stéréotypes de genre et de race en France.

2. Mise en place d'une expérience perceptive : stimuli, protocole et auditeurs

Nous avons présenté notre expérimentation comme étant un sondage sur la « première impression » produite par une voix. L'écran d'accueil que l'interface du test de perception présentait aux participants avant qu'ils ne puissent commencer à répondre aux questions comportait le texte suivant :

Bonjour et merci de participer à ce test dont le but est d'évaluer la première impression produite par la parole de jeunes candidats étudiants à un examen oral de personnalité.

Nota bene : la toute première impression n'est pas définitive, elle se forme dès les premières secondes d'écoute. Ainsi, vous allez écouter une dizaine de secondes de parole de chaque candidat/e. On vous conseille de répondre rapidement, sans trop réfléchir, pour garder cette première impression produite sur vous. ***Il n'y a pas de bonnes ou de mauvaises réponses à ce type de test !***

Figure 1 : Écran d'accueil du test de perception

2. 1. Stimuli

Les dix extraits sonores que nous avons utilisés comme stimuli dans le test de perception provenaient d'enregistrements que nous avons réalisés au cours de nos terrains de recherche ; ils sont produits par des locuteurs jeunes, de moins de 25 ans, dans des situations dans lesquelles était sollicitée une présentation de soi. Nous avons choisi d'utiliser des extraits d'une dizaine de secondes, des extraits donc relativement longs, parce que ceux-ci permettent une écoute globale, plus proche des pratiques d'écoute hors laboratoire. Ce choix nous semble important dans la mesure où il évite de focaliser artificiellement l'attention des auditeurs uniquement sur des détails de prononciation, déconnectés du sens, et d'activer des tâches cognitives totalement inhabituelles comme lors de l'écoute de séquences très courtes ou de logatomes (non-mots)³.

Nous avons ensuite manipulé certains de ces extraits sonores avec des procédés de resynthèse afin de créer, à l'intérieur du continuum des voix genrées possibles, un ensemble de voix androgynes et de voix féminines et masculines non-prototypiques.

Pour créer les voix androgynes, nous avons dans un premier temps resynthétisé les fréquences fondamentales et les fréquences de résonance d'un ensemble de voix selon le procédé expliqué par Arnold (2012). Ensuite, pour déterminer lesquelles des voix créées pouvaient être qualifiées d'« androgynes », nous avons fait un test de perception dans lequel nous avons demandé à dix auditeurs de catégoriser les voix entendues en tant que « voix de femmes » ou en tant que « voix d'homme ». Les voix ayant reçu un nombre égal de catégorisations en tant que « voix de femme » et en tant que « voix d'homme » ont été retenues comme « voix androgynes ».

C'est ce même procédé de resynthèse que nous avons utilisé pour créer les voix féminines et masculines non-prototypiques. Nous avons ainsi créé des voix de femmes plutôt graves et sombres⁴, perçues comme « moins féminines » que les voix originales à partir desquelles nous avons réalisé la resynthèse, et également des voix d'hommes plutôt aiguës et claires, perçues comme « moins masculines » que les voix originales.

Au final, les extraits sonores utilisés comportaient trois types de voix : des voix féminines et masculines prototypiques (les voix enregistrées dans leur version originale), des voix androgynes (obtenues par resynthèse) et des voix féminines et masculines non-prototypiques (obtenues par resynthèse).

2. 2. Protocole

Dans notre expérience, nous avons souhaité tester le degré d'approbation d'une

³ Voir Morange & Candea (2010) pour une discussion sur l'importance du choix des durées.

⁴ La fréquence fondamentale d'un son va déterminer s'il va être perçu comme grave ou aigu (hauteur) et les fréquences de résonance vont déterminer s'il va être perçu comme sombre ou clair (timbre).

série de six phrases affirmatives et évaluatives⁵, orientées positivement ou négativement, portant soit sur des qualités attitudinales du locuteur soit sur son élocution (tableau 1).

Évaluation selon une orientation positive
<i>Mathilde L. est sûre d'elle</i>
<i>Mathilde L. est sympathique</i>
<i>Mathilde L. est dynamique</i>
Évaluation selon une orientation négative
<i>Mathilde L. est hésitante</i>
<i>Mathilde L. est agressive</i>
<i>Mathilde L. parle d'une voix monotone et ennuyeuse</i>

Tableau 1 : Phrases soumises aux auditeurs

Ces phrases s'affichaient une à une à l'écran et étaient appariées à un extrait sonore. Chacune comportait un prénom féminin ou masculin, français ou maghrébin pour que les auditeurs associent les paroles entendues aux catégories identitaires « femme », « homme », « Français de souche », « d'origine arabe ».

Pour tester l'effet des stéréotypes nous avons conçu un protocole combinant des éléments de la technique du *locuteur masqué* (*matched guise* ; voir par exemple Lambert et *al.*, 1960) ainsi que des techniques d'*amorçage*⁶ (*priming* ; voir Meyer & Schvaneveldt, 1971) qui consistait, d'une part, à présenter un même extrait sonore associé visuellement à des identités différentes et, d'autre part, à présenter des extraits sonores différents associés à une même identité. En revanche, la longueur de nos stimuli ne permettait pas d'utiliser cette méthode au cours d'un seul et unique test de perception car les auditeurs auraient reconnu les extraits. Nous avons donc constitué deux groupes d'auditeurs indépendants de 55 personnes chacun.

Le tableau 2 récapitule les amorçages visuels utilisés pour chaque extrait sonore ciblé, selon le groupe.

⁵ Notre choix de phrases a été motivé par les résultats des nombreuses études menées sur les attitudes associées stéréotypiquement aux personnes catégorisées en tant que « femme », « homme » (p. ex. Williams & Bennett, 1975 ; Williams & Best, 1990), « blanc », « noir », « arabe » (p. ex. Dambrun & Guimond, 2004 ; Lacassagne et *al.*, 2001).

⁶ L'effet d'amorçage se définit comme l'influence de la présentation d'un premier stimulus (l'amorce) sur le traitement d'un stimulus consécutif (la cible).

Extraits sonores	Amorces Groupe 1 (55 auditeurs)	Amorces Groupe 2 (55 auditeurs)
Extrait 1	<i>Emilie G.</i>	<i>Emilie G.</i>
Extrait 2	<i>Martial T.</i>	<i>Martial T.</i>
Extrait 3	<i>Fouzia A.</i>	<i>Fouzia A.</i>
Extrait 4 (voix androgyne)	Mathieu L.	Mathilde L.
Extrait 5 (voix androgyne)	Fatiha A.-K.	Hassan A.-K.
Extrait 6 (voix masculine)	Paul H.	Samir H.
Extrait 7 (voix masculine)	Omar C.	Eric C.
Extrait 8 (voix féminine)	Céline Ph.	Zohra Kh.
Extrait 9 (voix féminine)	Ghenima Kh.	Constance Ph.
Extrait 10a (voix originale)	Farid B.	/
Extrait 10b (voix resynthétisée, moins masculine)	/	Farid B.
Extrait 11a (voix originale)	/	Vincent D.-T.
Extrait 11b (voix resynthétisée, moins masculine)	Vincent D.-T.	/
Extrait 12a (voix originale)	/	Nedjma M.
Extrait 12b (voix resynthétisée, moins féminine)	Nedjma M.	/
Extrait 13a (voix originale)	Irène S.-J.	/
Extrait 13b (voix resynthétisée, moins féminine)	/	Irène S.-J.

Tableau 2 : Jeu d'amorçages pour les deux groupes d'auditeurs

Les trois premiers extraits présentaient des amorçages identiques pour les deux groupes car ils servaient simplement d'entraînement. Les extraits 4 et 5, voix androgynes, ont été présentés d'un groupe à l'autre tantôt avec des prénoms féminins, tantôt avec des prénoms masculins. Les extraits 6 à 9, voix de femmes et d'hommes prototypiques, ont été présentés avec des prénoms à consonance tantôt française, tantôt maghrébine. Par exemple, tout en écoutant le même extrait sonore, le groupe 1 pensait écouter *Ghenima Kh.* tandis que le groupe 2 pensait écouter *Constance Ph.* Les extraits 10 à 13 comportaient les voix féminines et masculines non-prototypiques (10b, 11b, 12b et 13b) et les voix originales à partir desquelles les voix non-prototypiques avaient été resynthétisées (10a, 11a, 12a et 13a). À chaque groupe a été présentée uniquement une des deux versions : ou bien la version originale ou bien la version resynthétisée. Par exemple, le groupe 1 écoutait l'extrait 10a, associé à Farid B., tandis que le groupe 2 écoutait l'extrait 10b associé à la même étiquette Farid B.

L'interface que nous avons utilisée est reproduite dans la figure 2 ci-dessous. L'amorçage était supraliminal : le prénom suivi de l'initiale du nom était bien visible, en vert, à gauche, juste au-dessus du bouton d'écoute, et repris dans la phrase à évaluer placée au-dessus de l'échelle à six degrés (échelle allant de « pas du tout d'accord » à « tout à fait d'accord »).



Figure 2 : Capture d'écran de l'interface utilisée

Durant le test, chaque extrait audio était présenté six fois, associé au même prénom et à l'une de nos six phrases évaluatives du tableau 1. L'ordre de passage des extraits était généré de manière aléatoire par un script pour neutraliser les effets liés à l'ordre des réponses et à l'entraînement dans la tâche à effectuer.

2. 3. Auditeurs

Les 110 personnes qui ont répondu à nos tests ont été sollicitées de la même manière et réparties aléatoirement dans un des deux groupes. Il s'agissait essentiellement d'étudiants dans des disciplines littéraires ou linguistiques (dont une bonne part en formation continue) et de quelques enseignants dans ces mêmes disciplines. Une trentaine de personnes a passé le test devant un des enquêteurs, dans une petite salle à l'université, ce qui nous a permis de valider la bonne compréhension des consignes et l'ergonomie de l'interface. Les autres ont passé le test chez eux, sur internet, avec un casque ou des écouteurs. Les âges de nos sujets varient dans une plage de 19 à 64 ans (moyenne 32 ans) ; nous avons sollicité au total 110 personnes. Toutes ont appris le français comme première langue, sauf un très petit nombre de personnes d'origine étrangère qui vivent en France depuis plus de dix ans et ont un excellent niveau en français L2⁷.

3. Influence des stéréotypes de genre

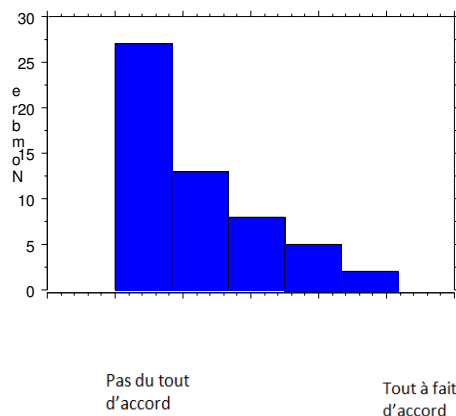
Nos résultats ont montré que certains couples d'amorces, comme par exemple le couple Fatiha et Hassan, avaient suscité des évaluations très similaires dans les deux groupes, tandis que d'autres, comme par exemple le couple Mathilde et Mathieu, avaient suscité des évaluations significativement différentes.

Au regard de la littérature sur les stéréotypes (voir *supra*), l'hypothèse que nous avons émise avant d'effectuer le test de perception était que les stéréotypes de genre induiraient des évaluations globalement plus négatives (plus sévères) pour les femmes et plus positives (plus indulgentes) pour les hommes. Cet effet ne

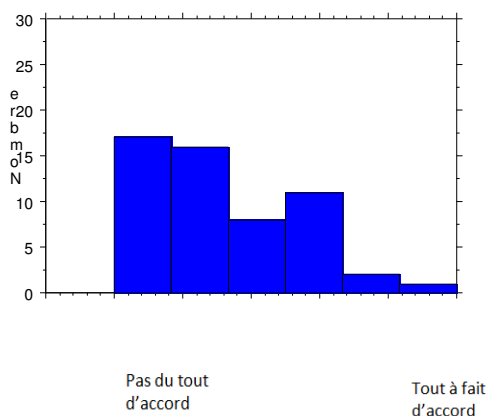
⁷ Le même test serait à effectuer avec des auditeurs vivant hors de France : pour l'instant, les résultats des auditeurs vivant hors de France ayant participé à notre expérience ont été écartés des calculs.

s'est pas produit. Au lieu de cela, nous avons trouvé des indices que nous avons interprétés comme une influence des représentations prototypiques des voix de femmes et des voix d'hommes.

Premièrement, nous avons remarqué qu'une voix androgyne avait été perçue différemment en fonction du prénom d'amorçage, Mathilde ou Mathieu. Lorsqu'elle a été présentée comme Mathilde, elle a suscité des jugements globalement plus positifs. Quand les auditrices et auditeurs pensaient qu'il s'agissait d'une locutrice, celle-ci était perçue comme plus dynamique, plus sympathique et plus sûre d'elle ; mais elle a également été perçue plus souvent comme agressive, comme le montre la Figure 3 qui compare l'effet des affichages « Mathieu L. » vs « Mathilde L. ».



**Figure 3 : Évaluations de la phrase
« Mathieu L. est agressif »**



**Figure 4 : Évaluations de la phrase
« Mathilde L. est agressive »**

La comparaison des deux séries de résultats nous a permis de constater une différence significative ($p < 0,05$; test U de Mann-Whitney⁸) en fonction du genre supposé de la personne écoutée. Ainsi, la phrase *Mathieu L. est agressif* a reçu un nombre largement supérieur d'évaluations « pas du tout d'accord » par rapport à la phrase *Mathilde L. est agressive*.

Les évaluations différentes en fonction du genre supposé des personnes écoutées peuvent être expliquées par plusieurs facteurs. Un premier facteur serait que les auditrices et auditeurs ont une représentation prototypique des voix de femmes et des voix d'hommes et qu'elles/ils jugent les stimuli *par rapport* à cette représentation. Ainsi, les voix androgynes présentées comme voix de femmes ont pu être perçues comme graves et sombres *par rapport* aux voix de femmes prototypiques, tandis que les voix présentées comme voix d'hommes ont pu être perçues comme aiguës et claires *par rapport* aux voix d'hommes prototypiques.

En outre, les auditrices et auditeurs ont pu associer aux voix entendues certaines attitudes liées à cette perception relative aux représentations prototypiques. Les voix présentées comme voix de femmes, perçues comme graves et sombres par rapport aux prototypes, ont pu évoquer des attitudes associées cognitivement

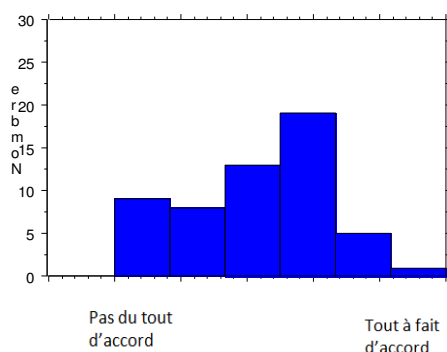
⁸ L'utilisation du test U de Mann-Whitney est motivée par la distribution non-paramétrique des réponses sur notre échelle à six degrés.

aux basses fréquences, tandis que les voix présentées comme voix d'hommes ont pu évoquer des attitudes associées aux fréquences élevées. Or, ces associations cognitives entre fréquences vocales et attitudes ont été largement documentées dans de nombreux travaux phonétiques (Apple et *al.*, 1979 ; Bolinger, 1964 ; Brown, 1974 ; Fairbanks & Pronovost, 1939 ; Ohala, 1984 ; Williams & Stevens, 1972). Selon ces travaux, les fréquences basses évoqueraient par exemple l'assertivité, l'autorité, la confiance en soi ou l'agressivité, et les fréquences élevées évoqueraient par exemple la soumission, la déférence ou le manque de confiance en soi. Cela nous incite à penser que les voix de femmes avec des fréquences basses (voix graves/sombres) sont globalement évaluées plus positivement que les voix d'hommes avec des fréquences élevées (voix aiguës/claires) grâce aux attitudes positives associées aux basses fréquences.

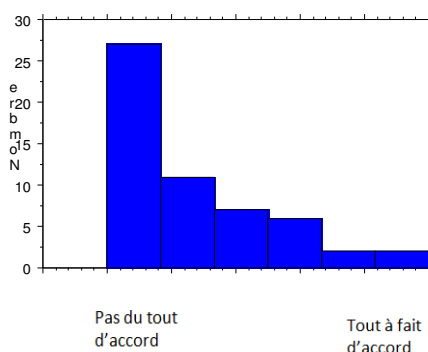
Un second facteur qui peut avoir joué un rôle est que les voix d'hommes aiguës sont souvent stéréotypiquement associées à l'homosexualité (Gaudio, 1994), alors que cela est moins systématiquement le cas avec les voix de femmes graves. Cela peut notamment être expliqué par le fait que les voix d'hommes gays sont plus souvent caricaturées, et donc stéréotypées, que celles des femmes lesbiennes (Moonwomon-Baird, 1997⁹). Les attitudes et stéréotypes négatifs à l'égard de l'homosexualité (voir par exemple Dayer, 2007) peuvent donc avoir provoqué des jugements plus négatifs lorsque les voix androgynes ont été présentées comme voix d'hommes.

En second lieu nous avons comparé les évaluations des voix féminines et masculines non-prototypiques – les voix que nous avons créées par resynthèse pour modifier leur degré de féminité/masculinité – avec les évaluations des voix prototypiques – les voix originales à partir desquelles nous avons créé les voix resynthétisées. Cette comparaison nous a permis de remarquer que les voix de femmes aiguës et claires étaient systématiquement jugées plus positivement que les voix de femmes graves et sombres. Le phénomène inverse a été observé pour les voix d'hommes. Ainsi, les voix qui se rapprochaient du prototype étaient systématiquement évaluées plus positivement que celles qui s'en écartaient. Cette tendance peut être illustrée à travers l'exemple du degré d'adhésion à la phrase *Vincent D.-T. est hésitant* associée à deux extraits sonores qui se différenciaient uniquement par le timbre de la voix : l'extrait 11a (voix originale de Vincent D.-T.) et l'extrait 11b (voix de Vincent D.-T. resynthétisée, « moins masculine »).

⁹ Moonwomon-Baird (1997) explique que les voix de femmes lesbiennes sont moins caricaturées que celles des hommes gays par le fait que les femmes sont de manière générale socialement moins visibles et audibles que les hommes.



**Figure 5 : Évaluations de la phrase
*Vincent D.-T. est hésitant avec voix
resynthétisée « moins masculine »***



**Figure 6 : Évaluations de la phrase
*Vincent D.-T. est hésitant avec voix
originale « plus masculine »***

La comparaison des évaluations dans les deux groupes nous a permis de constater une différence significative ($p < 0.0001$; test U de Mann-Whitney). Bien que les deux séries de réponses s'orientent globalement vers le désaccord avec la phrase *Vincent D.-T. est hésitant*, ce désaccord est très nettement marqué dans le groupe qui a entendu la voix originale, mais pas très catégorique dans celui qui a écouté la voix resynthétisée, « moins masculine ».

4. Influence des stéréotypes raciaux

Aucun des locuteurs de nos extraits n'a été classé majoritairement comme « pas du tout » *sympathique* ou comme « tout à fait » *agressif*, soit parce que nos extraits n'étaient pas assez clairement marqués pour indexer l'agressivité ou le caractère antipathique, soit parce que les auditeurs rechignent en général à donner des évaluations trop sévères à partir de l'écoute de si courts extraits. En outre, les réponses pour les catégories à orientation négative de notre protocole (*agressif, hésitant, voix monotone*) sont assez homogènes d'un groupe de juges à l'autre.

En revanche, si on examine comment nos catégories à orientation positive *dynamique, sympathique, sûr de soi* ont été choisies, nous avons pu mesurer un effet global mélioratif, statistiquement significatif, de l'indexation raciale par le prénom ($p < 0.04$; test U de Mann-Whitney). De manière inattendue, les extraits associés aux prénoms maghrébins sont globalement mieux évalués.

De même, la seule voix qui ait été globalement très appréciée selon nos six critères a été celle affichée comme *Farid B.* Il semblerait que l'amorçage par ce prénom ait « protégé » l'extrait sonore associé de toute évaluation négative, que ce soit pour le groupe 1 qui a écouté la version originale, proche du prototype de voix masculine, ou pour le groupe 2 qui a écouté la version resynthétisée de manière à ce qu'elle soit « moins masculine ». En cela, cet extrait associé au prénom Farid se distingue des autres voix resynthétisées, affichées comme *Vincent D.-T., Irène J.-S. et Nedjma M.*, pour lesquelles la modification introduisant un écart entre l'attente induite par la catégorisation masculine ou féminine du prénom et ce qui était entendu, avait entraîné une dégradation dans les évaluations (voir *supra*).

En outre, si on observe les évaluations obtenues pour les catégories *dynamique* et *sûr/e de soi*, il apparaît que les extraits associés aux prénoms féminins (français ou maghrébins) ainsi que ceux associés aux prénoms masculins français obtiennent des appréciations dispersées dans les trois zones de notre échelle à 6 degrés : négative (0 ou 1), mitigée (2 ou 3) ou positive (4 ou 5). En revanche, les appréciations obtenues pour les extraits affichés avec des prénoms masculins maghrébins sont davantage concentrées dans le pôle positif (perçus donc comme plutôt dynamiques ou sûrs de soi). La seule exception à cette tendance est l'extrait sonore associé à *Samir H.* qui a reçu des appréciations majoritairement mitigées, mais l'effet positif du prénom à consonance maghrébine est tout de même présent dans la mesure où le même extrait affiché comme *Paul H.* a reçu des appréciations plus négatives. Les figures 7 et 8 ci-dessous montrent le déplacement des évaluations vers le pôle de l'accord avec la phrase *X est sympathique* lorsqu'elle a été associée à *Samir H.* par rapport au groupe où elle était associée à *Paul H.* ($p < 0.01$; test U Mann-Whitney,).

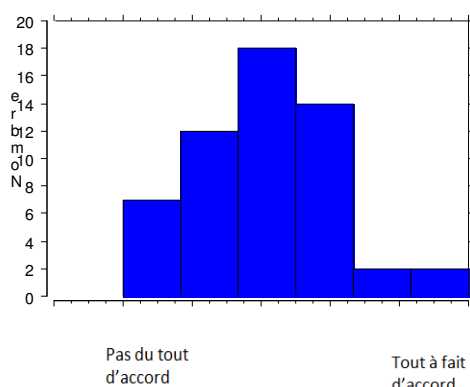


Figure 7 : Évaluations de la phrase *Paul H. est sympathique*

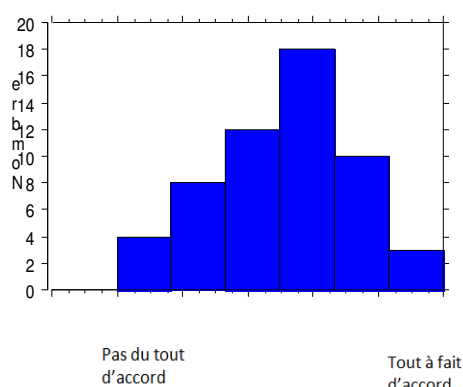


Figure 8 : Évaluations de la phrase *Samir H. est sympathique*

Or, ce qui peut apparaître comme une préférence pour une voix masculine associée à un prénom maghrébin, semble contradictoire avec les études actuelles (par exemple Lorcerie, 2011) qui montrent que les hommes et les garçons catégorisés comme « arabes » subissent plus de discrimination de la part des groupes dominants que les femmes et les filles de ce même groupe. Comment interpréter dès lors nos résultats inversés ?

L'hypothèse que nous avançons pour les résultats porte plutôt sur la stratégie adoptée pour répondre à notre test. Nous pensons que les participants étaient conscients de l'existence dans la société des discours stigmatisants et xénophobes à l'encontre des personnes immigrées ou catégorisées comme « issues de l'immigration arabe » et ont eu soin de s'en démarquer en adoptant une stratégie inverse, qui a abouti à favoriser les extraits affichés avec des prénoms maghrébins¹⁰.

Certains échanges avec nos auditeurs et auditrices ont corroboré notre

¹⁰ Une observation similaire a été faite par Brouard & Tiberj (2005) qui ont mené un test portant sur la sélection d'un candidat (Ahmed vs Vincent) ; la sélection était uniquement déclarative, en situation de test, et les sujets avaient tendance à accorder un avantage à « Ahmed » lorsque des questions de justice sociale étaient évoquées durant le test.

interprétation des résultats par un mécanisme d'autocensure et de compensation. A titre d'exemple, nous pouvons citer un e-mail reçu en retour d'une ancienne étudiante de master qui avait accepté de participer à notre expérience. Elle a écrit : « Durant l'écoute, j'ai pensé à différentes choses comme : lui il a un accent de banlieue, mais il n'est pas agressif, c'est sa façon "normale" de parler ! » La négation polémique « pas agressif », s'oppose à une appréciation qu'elle reconnaît comme pouvant être induite par l'extrait qu'elle venait d'écouter, et qu'elle souhaite consciemment inhiber.

Cependant, si notre proposition d'interprétation est juste, elle laisse sans réponse un grand nombre de questions qui échappent à notre protocole. En effet, on peut se demander si, ayant intégré qu'il n'est pas acceptable d'appliquer des préjugés raciaux lorsqu'on évalue quelqu'un, les auditeurs ont simplement censuré leurs réponses (« je ne dois pas dire que je pense X »), ou bien s'ils sont tout à fait conscients de ce danger et ils font tout pour l'écarter (« je ne dois pas penser X »). On peut même se demander si une partie des auditeurs n'est pas consciente de l'exploitation qui peut être faite de leurs réponses et n'adopte pas une stratégie de masquage d'attitude discriminatoire pour que celle-ci ne puisse pas être mesurée par ce type de test¹¹. Un petit sondage *a posteriori* sur les opinions politiques de nos auditeurs nous laisse penser que celles et ceux qui se définissent politiquement à gauche étaient surreprésentés, tandis que celles et ceux qui se définissent comme sympathisants de la droite ou de l'extrême droite n'y sont pas représentés (ou que ceux-ci souhaitaient ne pas être identifiés comme tels). Pour autant, il nous semble qu'il aurait été impossible d'opérer une sélection préalable des auditeurs selon leurs opinions politiques déclarées, car poser une telle question n'aurait pu que renforcer le biais en attirant encore plus clairement et explicitement leur attention sur l'exploitation politique des réponses.

Si notre hypothèse est juste, et s'il est vrai qu'il existerait un écart important de niveau de conscience entre les stéréotypes de genre et les stéréotypes de race, aujourd'hui en France, cela impliquerait une difficulté méthodologique de rendre opérationnel, dans des études portant sur la perception d'autrui, le concept d'intersectionnalité (Crenshaw, 1989) que l'on pourrait résumer très brièvement par l'idée que les oppressions de genre et de race ne s'additionnent pas mais interagissent entre elles de manière complexe.

5. Interroger l'usage des protocoles quantitatifs pour l'étude des stéréotypes

Notre étude s'inscrit dans la tradition de réflexion sur les méthodes expérimentales. Nous sommes partis du postulat qu'une expérimentation scientifique est d'abord une interaction entre humains et, dans ce sens, un

¹¹ Une étudiante en master qui a mené un test de perception pour un devoir en sociolinguistique sous la direction de M. Candea rapporte dans son compte rendu une attitude interprétable en ce sens. Elle note, au sujet d'un participant qui avait fait des études en sciences sociales : « A l'écoute de l'extrait 8, il dit en riant « En fait depuis le début elle veut m'avoir, c'est que des Tourangeaux, et celui qui parle bien, il est Algérien » [...] Toutes ses réponses passeront par ce filtre d'appréhension du piège de l'exercice. » Beaucoup de gens pensent souvent à juste titre qu'ils se feront piéger par un test à visée sociologique et réagissent en fonction de cela.

espace intersubjectif. En outre, les attentes des individus sollicités pour une telle tâche de jugement dans une culture à un moment donné (en l'occurrence en France en 2012, au moment où l'expérience a été menée) dépendent de plusieurs variables : l'expérience des juges, leur formation, leurs lectures, leur familiarité avec les enquêtes d'opinion et avec les thèmes de recherche actuels, ou même spécifiquement avec les thèmes de recherche ou les objectifs possibles des enquêteurs si les juges se forment *a priori* une idée à ce sujet, etc. Contrôler ou neutraliser ces variables nous semble utopique, et il serait vain de contourner cet obstacle ou de l'ignorer.

Certes, le mécanisme d'autocensure est connu depuis longtemps par les chercheurs en psychologie sociale qui ont imaginé différents protocoles pour « diminuer les possibilités de contrôle volontaire des participants sur leurs réponses » (Guimond, 2006)¹². Une des techniques possibles préconise les amorçages subliminaux, comme chez Dambrun et Guimond (2004) et Guimond (2006). Ces chercheurs font l'hypothèse que si la perception subliminale à l'écran du mot *arabe* (versus *français*) rend plus rapide l'identification de mots négatifs comme *violent*, *agressif*, *voleur*, *ignorant* affichés dans une série de logatomes, cela permettrait de mesurer les stéréotypes négatifs des participants qui n'auraient probablement pas été dévoilés dans un questionnaire explicite.

Cependant, cette technique ne réglerait pas tous les problèmes que nous soulevons dans notre étude : qu'est-ce qui garantit que le fait que la perception subliminale du mot *arabe* active plus rapidement l'identification du mot *agressif* ou *voleur* serait une mesure des préjugés du participant ? On pourrait en effet avancer une hypothèse différente : l'activation plus rapide de ces identifications pourrait mesurer la fréquence des stéréotypes en question dans les discours auxquels les participants ont été exposés. Si on tient compte des analyses de Butler (1997) pour qui la force d'une injure langagière vient de son itération, il n'est pas impossible que la personne qui perçoit *arabe* de manière subliminale active les stéréotypes en raison de leur fréquence dans les discours et non obligatoirement en raison de sa propre vision stigmatisante. Autrement dit, on peut très bien imaginer que quelqu'un qui adhère fortement à l'idée que « non, les Arabes ne sont pas plus agressifs que les Français » active tout autant le mot *agressif* après l'amorce *arabe* dans une tâche cognitive décontextualisée, en raison de sa connaissance de la prégnance de ce stéréotype.

L'interprétation que nous proposons de nos résultats interroge, par ailleurs, de manière plus générale, les approches quantitatives des pratiques sociales qui écartent du protocole toute prise en compte de facteurs comme les représentations liées aux genres ou aux races.

Nos hypothèses nous amènent par conséquent à formuler des réserves notamment sur la portée des résultats d'expériences perceptives menées en utilisant des voix féminines ou des voix masculines, sans aucune interrogation sur les représentations genrées des juges. Si l'on revient par exemple à l'étude très connue de Rubin (1992), l'on peut constater que celle-ci a été menée uniquement sur deux photos d'enseignantes féminines (une « asiatique », une « caucasienne »). Cependant, à aucun moment le genre n'a été questionné. Les

¹² On peut se rapporter à Dambrun et Guimond (2003) pour un état de l'art sur ces questions.

photos n'étant pas incluses dans la publication, il est impossible de savoir, par exemple, quel pouvait être l'écart entre les photos présentées et les prototypes de féminité. Il en est de même pour Niedzielski (1999) qui a fait écouter uniquement des voix de locuteurs féminins à ses auditeurs. Nous pourrions également évoquer une étude récente menée en France par Boula de Mareuil et Lehka-Lemarchand (2012) qui porte sur la perception de l'accent dit « de banlieue » (au sens de périphérie urbaine pluriethnique et populaire). Les auteurs y utilisent uniquement des voix de jeunes hommes pour analyser la perception sans y interroger les représentations genrées et racialisées liées à la banlieue, et sans donner la moindre indication sur le timbre de ces voix ou sur leur proximité avec les prototypes de voix masculines. Les exemples de ce type peuvent être multipliés à l'infini dans le champ des études sur la perception de la parole : nos interrogations ne remettent pas en question les résultats obtenus ainsi, mais l'interprétation trop généralisante qui pourrait en être faite.

6. Conclusions

Pour conclure, notre étude visait à montrer l'intérêt que peut avoir un questionnement de l'influence des catégories de genre et de race dans les tests de perception qui utilisent des échantillons de voix humaines. Notre démarche a été motivée par le constat que cette influence était bien souvent totalement ignorée dans les protocoles de test, soit par méconnaissance des travaux précédents, soit par omission délibérée de la part de chercheurs qui se déclarent parfois mal à l'aise avec ces sujets, probablement en lien avec l'absence d'une tradition de recherche suffisamment conséquente en France.

Au final, notre conclusion est double car elle comporte un volet spécifique sur la notoriété, à l'heure actuelle, des stéréotypes liés au genre et à la race et également un volet plus général sur les méthodes expérimentales d'investigation.

D'un côté, notre étude a abouti à quelques résultats inattendus sur les stratégies pour évaluer des voix humaines présentées avec différents prénoms féminins ou masculins, français ou arabes. L'analyse de ces évaluations nous conduit à penser que les stéréotypes de genre sont moins connus par le public ou moins évités, moins dénoncés explicitement, que les stéréotypes raciaux. Il est possible que les gens aient un niveau de conscience supérieur de l'utilisation qui peut être faite de leurs réponses en matière d'analyse des discriminations raciales qu'en matière de discriminations liées au genre. Si c'est le cas, cela voudrait dire qu'il serait plus problématique de distinguer entre l'absence de stéréotypes raciaux négatifs et les stratégies de masquage de ces stéréotypes, et que ce problème, en tout cas pour le moment en France, se poserait moins sur les stéréotypes de genre qui ne bénéficient pas de la même notoriété.

En outre, notre enquête laisse penser que des méthodes d'investigation comme celles utilisées par exemple dans la célèbre étude de Katz et Braly (1933), où il fallait associer des qualificatifs à des catégories raciales, ne fonctionneraient plus aujourd'hui. Cela ne s'explique pas seulement par l'autocensure des auditeurs ou par un supposé recul des stéréotypes racistes, mais aussi par le fait que les participants à une expérience perceptive ont à présent une culture plus riche sur

les sujets abordés par les études à visée sociologique, et développent une méfiance structurée à leur égard. La complexité de ces sujets plaide d'une part pour des approches qualitatives permettant de mieux appréhender l'interaction entre chercheurs et enquêtés et d'autre part pour la prise en compte plus systématique, dans les jeux de stimuli vocaux utilisés dans les tests de perception en général, des représentations genrées ou raciales que ceux-ci peuvent activer de manière implicite.

Loin de vouloir rejeter toute approche expérimentale et quantitative des pratiques sociales nous plaidons pour une combinaison systématique de techniques d'analyse quantitatives et qualitatives. L'utilisation de *méthodes mixtes* (Creswell & Clark, 2006 ; Tashakkori & Teddlie, 2002) nous semble être devenue indispensable. Ce n'est qu'en prenant constamment « tout ce que l'on croit savoir pour hypothèse » (Heller, 2002) que nous pourrions faire en sorte que la recherche en sociolinguistique continue d'être « engagée, critique et réflexive » (*ibid.*).

Bibliographie

- Apple, W., Streeter, L. A. & Krauss, R. M. (1979), Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37(5), p.715–727.
- Arnold, A. (2012), Le rôle de la fréquence fondamentale et des fréquences de résonance dans la perception du genre. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, 28 [en ligne].
- Bolinger, D. (1964), Intonation as a Universal, in H. G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists* The Hague, Mouton, p. 833–848.
- Boula de Mareüil, P. & Lehka-Lemarchand, I. (2011), Can a prosodic pattern induce/reduce the perception of a lower-class suburban accent in French ? *Proceedings of the XVIIth International Congress of Phonetic Sciences (ICPhS)*. Hong Kong.
- Brouard, S. & Tiberj, V. (2005), *Français comme les autres ? Enquête sur les citoyens d'origine maghrébine, africaine et turque*. Paris, Presses de Sciences Po.
- Brown, B. L. (1974), Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *The Journal of the Acoustical Society of America*, 55, 313-318.
- Bruner, J. S. (1957), On perceptual readiness. *Psychological Review*, 64(2), p.123–152.
- Butler, J. (1990), *Gender trouble: feminism and the subversion of identity*, New York, Routledge.
- Butler, J. (1993), *Bodies that matter: on the discursive limits of 'sex'*. New York, Routledge.
- Butler, J. (1997), *Excitable speech: a politics of the performative*. New York, Routledge.
- Butler, J. (2004), *Undoing gender*. New York, Routledge.

- Condry, J. & Condry, S. (1976), Sex Differences: A Study of the Eye of the Beholder, *Child Development*, 47(3), p.812–819.
- Crenshaw, K. (1989), Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics, *University of Chicago Legal Forum*, 139-167.
- Creswell, J. W. & Clark, D. V. L. P. (2006), *Designing and Conducting Mixed Methods Research (1st ed.)*, Thousand Oaks, Sage.
- Dambrun, M. & Guimond, S. (2003), Les mesures implicites et explicites des préjugés et leur relation : développements récents et perspectives théoriques. *Les Cahiers internationaux de psychologie sociale*, 57, p.52–73.
- Dambrun, M. & Guimond, S. (2004), Implicit and explicit measures of prejudice and stereotyping: do they assess the same underlying knowledge structure? *European Journal of Social Psychology*, 34(6), p.663–676.
- Dayer, C. (2007), Stéréotypes et homosexualité: entre attribution et démarcation, dans H. Boyer (éd.), *Stéréotypage, stéréotypes: Identités*, Paris, L'Harmattan, p.71–84.
- Dorlin, E. (2006), *La matrice de la race : Généalogie sexuelle et coloniale de la nation française*, Paris, La Découverte.
- Dorlin, E. (2008), *Sexe, genre et sexualités*. Paris, PUF.
- Fairbanks, G. & Pronovost, W. (1939), An experimental study of the pitch characteristics of the voice during the expression of emotion. *Communication Monographs*, 6, p.87–104.
- Gaudio, R. P. (1994), Sounding Gay: Pitch Properties in the Speech of Gay and Straight Men. *American Speech*, 69(1), p.30–57.
- Geis, F. L. (1993), Self-fulfilling prophecies: A social psychological view of gender, in A. E. Beall R. J. Sternberg (ed.), *The psychology of gender*, New York, Guilford Press, p.9–54.
- Gregory, R. L. (1980), Perceptions as Hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 290(1038), p.181–197.
- Guillaumin, C. (1992), *Sexe, race et pratique du pouvoir: L'idée de nature*. Paris, Côté-Femmes.
- Guimond, S. (2006), La fonction sociale des préjugés ethniques. *Cahiers de l'Urmis*, 10-11 [on line].
- Hall, K. (2005), Intertextual Sexuality: Parodies of Class, Identity, and Desire in Liminal Delhi. *Journal of Linguistic Anthropology*, 15(1), p.125–144.
- Heller, M. (2002), *Éléments d'une sociolinguistique critique*. Paris, Didier.
- Hilton, J. L. vo. H. & Von Hippel, W. von. (1996), Stereotypes. *Annual Review of Psychology*, 47(1), p.237-271.
- Hippel, W. von, Sekaquaptewa, D. & Vargas, P. (1995), On The Role Of Encoding Processes In Stereotype Maintenance. *Advances in Experimental Social Psychology* 27, p.177–254.
- Johnson, K., Strand, E. A. & D'Imperio, M. (1999), Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), p.359–384.
- Katz, D. & Braly, K. (1933), Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, 28(3), p.280–290.
- Lacassagne, M.-F., Salès-Wuillemin, E., Castel, P. & Jébrane, A. (2001), La catégorisation d'un exogroupe. *Papers on Social Representations - Textes sur les représentations sociales*, 10, p.7.1–7.11.

- Lambert, W. E., Hodgson, R. C., Gardner, R. C. & Fillenbaum, S. (1960), Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, 60(1), p.44–51.
- Lopez, I. F. H. (1994), The Social Construction of Race: Some Observations on Illusion, Fabrication, and Choice. *Harvard Civil Rights-Civil Liberties Law Review*, 29, p.1-62.
- Lorcerie, F. (2011), École et ethnicité en France : pour une approche systémique contextualisée. *SociologieS* [en ligne].
- Meyer, D. E., & Schvaneveldt, R. W. (1971), Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), p.227-234.
- Monneyron, F. & Siary, G. (2012), *L'idée de race. Histoire d'une fiction*, Paris, Berg International.
- Moonwomon-Baird, B. (1997), Towards a Study of Lesbian Speech, in A. Livia & K. Hall (eds.), *Queerly phrased: language, gender, and sexuality*, Oxford University Press, p.202–213.
- Morange, S. & Candea, M. (2010), Aux frontières de l'écoute. Durée des échantillons et choix des auditeurs : deux variables déterminantes dans la construction des tests de perception, dans D. Delomier & M.A. Morel (éds.) *Frontières, du linguistique au sémiotique*, Limoges, Lambert-Lucas, p.79-96.
- Niedzielski, N. (1999), The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, 18(1), p.62–85.
- Ohala, J. (1984), An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica*, 41, p.1–16.
- Rock, I. (1983), *The Logic of Perception*. MIT Press.
- Rubin, D. L. (1992), Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), p.511–531.
- Strand, E. A. (1999), Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology*, 18(1), p.86 –100.
- Strand, E. A. (2000), *Gender Stereotype Effects in Speech Processing*. The Ohio State University.
- Strand, E. A. & Johnson, K. (1996), Gradient and visual speaker normalization in the perception of fricatives, in D. Gibbon (ed.), *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, Bielefeld, Mouton, p.14–26.
- Tashakkori, A. & Teddlie, C. (2002), *Handbook of Mixed Methods in Social & Behavioral Research*, Thousand Oaks, Sage.
- Williams, C. E. & Stevens, K. N. (1972), Emotions and speech: some acoustical correlates. *The Journal of the Acoustical Society of America*, 52(4), p.1238–1250.
- Williams, J. & Bennett, S. (1975), The definition of sex stereotypes via the adjective check list. *Sex Roles*, 1(4), p.327-337.
- Williams, J. E. & Best, D. L. (1990), *Measuring sex stereotypes: a multinational study*, Thousand Oaks, Sage.