Clarification et correction d'indices segmentaux : une étude pilote sur les consonnes occlusives du français

Maëva Garnier, Marion Dohen, Louis Buttiaux, Silvain Gerber
Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France
* Institute of Engineering Univ. Grenoble Alpes
maeva.garnier@gipsa-lab.grenoble-inp.fr

RESUME

Cette étude aborde la question du trait de voisement et de la place d'articulation des consonnes occlusives du Français à l'aide d'un nouveau paradigme de correction segmentale, de façon à tester 1) quels indices de ces consonnes des locuteurs renforcent lorsqu'ils ont été mal compris par leur interlocuteur, et 2) si le renforcement de ces indices est toujours le même ou dépend de la consonne entendue à la place. Une locutrice jouait avec l'expérimentatrice à un jeu conçu pour simuler des situations d'incompréhension assez naturelles. La locutrice a montré plusieurs modifications de ses consonnes occlusives en parole claire et en situation de correction d'un mot mal compris (durée de la phase d'occlusion, VOT, intensité du bruit, …). En revanche, quasiment aucun descripteur n'a été modifié de façon différente en fonction du type d'erreur perceptive commise par l'interlocutrice (portant sur le trait de voisement ou sur le lieu d'articulation).

ABSTRACT

Clarification and correction of segment cues: a pilot study on French stop consonants.
This study deals with the question of the voicing feature and the articulation place of French stop consonants, using a new paradigm of segment correction. It aims at testing 1) which acoustic cues speakers enhance when they are misunderstood by an interlocutor, and 2) whether this enhancement is always the same or depends on the misperceived consonant. One speaker played with the experimenter a game designed to simulate natural situations of misunderstanding. The speakers showed several modifications of her stop consonants in clear speech and in situations of segment corrections (duration of the occlusion phase, VOT, burst intensity, …). However, hardly no descriptor was modified in a different way, depending on the perceptual error made by the interlocutor (on the voicing feature or on the articulation place).

MOTS-CLÉS: Interaction face-à-face; Consonnes occlusives; Traits; Parole claire; Multimodalité
KEYWORDS: Face-to-face interaction; Stop consonants; Features; Clear speech; Multimodality

1 Introduction

Sur quels indices acoustiques nous basons-nous en tant que locuteur ou auditeur, pour distinguer les différentes catégories de consonnes occlusives (/p/, /b/, /t/, /d/, /k/, /g/ pour les consonnes occlusives orales du Français) ? Cette distinction est-elle particulièrement claire lorsque ces indices sont « prototypés », prenant des valeurs très spécifiques, ou bien lorsque ces indices sont « contrastés », prenant des valeurs les plus distinctes possibles entre deux catégories phonologiques ? Ces questions, loin d'être nouvelles, ont été traitées extensivement ces 50 dernières années par de nombreuses études phonétiques en production et perception de la parole.

Les études sur la production des consonnes occlusives ont décrit des différences physiologiques et acoustiques entre différentes catégories de consonnes. En Français, l'opposition entre des consonnes occlusives voisées et non voisées se base littéralement sur la présence vs. l'absence de vibration des plis vocaux, et par conséquent sur la présence vs. l'absence d'énergie périodique audible pendant la phase d'occlusion. En Français, les segments [b d g] montrent un VOT (Voice Onset Time) négatif tandis que les segments [p t k] sont caractérisés par un VOT positif. Cette principale différence de production a plusieurs autres conséquences, directes ou directes, sur la longueur de la transition du premier formant (F1) (Liberman et al. 1954), la fréquence initiale de sa réapparition (Lisker et al. 1978), la fréquence fondamentale (f0) à la reprise de voisement, après le relâchement de l'occlusion (Ohde 1984), la durée de la voyelle précédente (Delattre 1962; Abdelli-Beruh 2004) ou encore la durée de la phase d'occlusion (Ohala et Riordan 1979; Abdelli-Beruh 2004).

Le lieu d'articulation d'une consonne occlusive configure le volume des cavités avant et arrière du conduit oral. La cavité avant joue un rôle déterminant sur la fréquence de la 2ème et 3ème résonance du conduit vocal au moment du relâchement de l'occlusion et sur leur variation jusqu'à la voyelle suivante. Elle détermine également l'allure de l'enveloppe spectrale du bruit créé au relâchement de l'occlusion. Pour les occlusives labiales [p b], l'intensité du bruit de plosion est faible, avec un spectre diffus-descendant (Lousada et al. 2012; Forrest et al. 1988) et une transition du deuxième formant vocalique (F2) montante devant la plupart des voyelles. Pour les occlusives apico-alvéolaires (ou dentales, par abus de langage) [t d], la cavité avant du conduit vocal est étroite, contribuant à un bruit de plosion d'intensité moyenne, avec un spectre diffus-montant et une transition de F2 descendante devant des voyelles centrales et postérieures, et légèrement montante devant des voyelles antérieures. Enfin pour les occlusives vélaires (ou palatales, par abus de langage) [k g], l'occlusion fait converger le 2ème et le 3ème formant vocalique (F2 et F3), se traduisant par un bruit de plosion de forte intensité, au spectre compact. Stevens et Blumstein (1978) ont soutenu l'idée que ces caractéristiques spectrales soient relativement invariantes et spécifiques à chaque lieu d'articulation. Ces principales différences spectrales sont également accompagnées d'autres variations acoustiques : Plus l'occlusion du conduit vocal est postérieure, et plus le VOT est long (Cho et Ladefoged 1999 ; Lisker et Abramson 1964), la transition de F1 courte et sa fréquence de réapparition haute (Summerfield et Haggard 1977).

De nombreuses études perceptives ont confirmé ces contrastes observés en production, en montrant dans quelle mesure la variation isolée ou combinée de ces différents traits acoustiques affecte la catégorisation perceptive d'un son plosif. Ainsi, il a été confirmé que la perception du trait de voisement était en effet affectée par la variation du VOT (Serniclaes 1984; Williams 1977), de l'intensité du bruit de plosion (Repp 1978; 1983; Williams 1977), de la valeur initiale de f0 (Haggard 1981) et de F1 après le relâchement de l'occlusion (Liberman et al.,1954 ; Summerfield et Haggard 1977), par la longueur de la transition de F1 (Stevens et Klatt 1974), la durée de la phase d'occlusion et de la voyelle précédente (Lisker 1957; Crystal et House 1988). De même, il a été confirmé que la perception du trait de place d'articulation était en effet affectée par les variations spectrales du bruit de plosion (Gravel 1983), la direction des transitions du 2ème et du 3ème formant (Delattre et al. 1955; Harris et al. 1958), la longueur du VOT et des transitions formantiques (Lisker et Abramson 1970; Lisker 1975; Miller 1981).

Il reste néanmoins un certain nombre de questions concernant ces indices, en particulier la question de leur hiérarchie et de leur degré d'importance, possiblement variable en fonction des situations de communication, en particulier dans des situations où les indices principaux sont ambigus ou altérés (parole chuchotée, environnement bruyant) (Winn et al. 2013). C'est pourquoi nous proposons dans ce projet une nouvelle approche complémentaire de ces questions en examinant, au travers d'une expérience de production de parole en interaction face-à-face, 1) quels indices de ces consonnes des locuteurs renforcent lorsqu'ils ont été mal compris par leur interlocuteur, et 2) si le renforcement de ces indices est toujours le même ou dépend de la consonne entendue à la place.

2 Matériel et méthodes

Une locutrice de 24 ans, de langue maternelle française, sans trouble auditif ni de la parole a participé à cette expérience. Cette locutrice était naïve vis-à-vis de l'objet de l'expérience et ne connaissait aucun des expérimentateurs.

L'expérience consistait en un jeu interactif, dans lequel la locutrice devait donner des consignes à une interlocutrice (auteure MD) pour avancer dans un labyrinthe. Chacune des deux interlocutrices disposait devant elle, sur un écran, d'une grille de 5x5 cases sur laquelle étaient représentées la case de départ (verte) et la case d'arrivée (rouge) d'un chemin, ainsi que des pseudo-mots sur les autres cases (cf. Figure 1). Sur la grille de la locutrice étaient également représentés les murs d'un labyrinthe, déterminant le chemin pour aller de la case de départ à la case d'arrivée. La tâche de la locutrice consistait à décrire pas à pas à l'expérimentatrice les différentes étapes du chemin (déplacements d'une case en horizontal ou en vertical). L'expérimentatrice devait tracer sur la grille une flèche représentant le déplacement compris, tout en continuant de parler naturellement.
Le but du labyrinthe était de susciter des situations assez naturelles d'incompréhension. A chaque étape, le déplacement pouvait en effet aboutir à des cases présentant des mots très proches auditivement. De ce fait, l'expérimentatrice pouvait parfois faire semblant, de façon assez crédible et naturelle, de ne pas avoir compris la consigne. Un exemple d'interaction est donné ci-dessous.

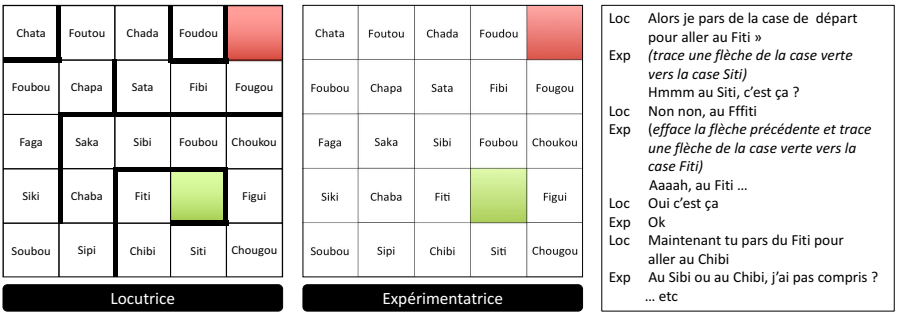

FIGURE 1: Exemple de grilles de jeu dont disposait la locutrice (à gauche) et l'expérimentatrice (à droite). La locutrice devait donner des consignes à l'expérimentatrice pour tracer le chemin partant de l'entrée du labyrinthe (case verte) à sa sortie (case rouge).

		C2					
		p	t	k	b	d	g
V	a	ʃapa	ʃata	ʃaka	ʃaba	ʃada	ʃaga
	i	sipi	siti	siki	sibi	sidi	sigi
	u	fupu	futu	fuku	fubu	fudu	fugu

TABLE 1 : Liste des 18 mots-cibles de type C_1VC_2V étudiés dans cette expérience.

Le jeu interactif permettait de faire produire à notre locutrice 18 mots cibles de type C_1VC_2V où C_2 est une consonne occlusive du Français (/p/, /t/, /k/, /b/, /d/ ou /g/) en contexte vocalique /a/, /i/ ou /u/ (cf. Table1). Ces mots ont été choisis de façon à avoir des ensembles de 4 mots différant par un seul segment (ici une consonne occlusive), présentant soit le même trait de voisement, mais un lieu d'articulation différent (ʃapa vs. ʃata et ʃaka), soit le même lieu d'articulation mais une opposition de voisement (ʃapa vs. ʃaba). De tels quadruplés étaient difficilement trouvables dans le lexique Français. C'est pourquoi nous avons fait le choix de pseudo-mots, traités comme des noms propres dans le contexte du jeu. Aucun d'entre eux ne correspondait à un mot réel, si bien que nous pouvons considérer leur familiarité comme identique. Pour le naturel du jeu, le corpus comportait également

36 autres pseudo-mots se distinguant de nos 18 mots cibles par la consonne initiale (C1) ou par la voyelle (V). Ces mots n'étaient que des « fillers » et n'ont pas été analysés.

Les deux interlocutrices interagissaient dans deux pièces séparées, à l'aide d'un système de visioconférence. Les deux interlocutrices étaient assises et portaient un micro-casque (AKG HSD 171), relié à une carte son (RME Fireface 800) permettant en même temps d'acquérir les deux signaux audio (à fe=44.1 kHz) et de renvoyer dans le casque de chaque participante un retour calibré de sa propre voix et celle de sa partenaire. La locutrice était filmée avec une caméra (25 images/s) située face à elle, derrière un écran/prompteur sur lequel était diffusé la vidéo de l'expérimentatrice. L'expérimentatrice était également filmée de face, grâce à une webcam miniature positionnée au centre de l'écran placé devant elle, diffusant la vidéo de la locutrice. Ce dispositif permettait que les interlocutrices interagissent en se regardant dans les yeux. En plus de cet écran principal, la locutrice disposait également d'un écran sur la gauche, représentant sa grille et d'un écran sur la droite, représentant la grille de l'expérimentatrice, et diffusant en temps réel les flèches tracées par l'expérimentatrice sur une tablette graphique.

L'expérience comportait 23 parties du jeu, avec une grille différente dans chaque partie. Durant les 5 premières parties, les deux interlocutrices interagissaient en condition de visioconférence « normale », sans perturbation (condition SP). Ces parties se déroulaient facilement, sans aucune incompréhension, et servaient de référence à la production de parole conversationnelle par la locutrice. Les 5 parties permettaient d'enregistrer 3 occurrences de chacun des 18 mots-cibles en parole conversationnelle.
Durant les 18 parties suivantes, la locutrice était informée qu'une perturbation était introduite sur le canal audio du système de visioconférence, de telle façon que l'expérimentatrice allait avoir des difficultés à la comprendre. En réalité, aucune perturbation n'était introduite mais l'expérimentatrice faisait effectivement semblant de mal entendre, amenant la locutrice à parler de façon plus claire (condition APN). L'expérimentatrice disposait d'un scénario très contrôlé lui indiquant, étape par étape de chaque grille, ce qu'elle devait faire semblant d'avoir compris. Ainsi, assez régulièrement (mais non systématiquement), l'expérimentatrice faisait semblant de se tromper, poussant la locutrice à répéter sa consigne, et donc le mot cible, en le corrigeant par rapport au mot mal compris par l'expérimentatrice (condition APC) (cf. exemple d'interaction Figure 1). De façon contrôlée, l'incompréhension (et donc la correction qui en découlait) portait soit sur le trait de voisement (1/3 du temps), soit sur le lieu d'articulation (1/3 du temps pour chacun des 2 autres lieux d'articulation que celui de la consonne cible, par exemple lieu dental ou palatal pour une consonne labiale, et réciproquement). Les 18 parties permettaient ainsi d'enregistrer 6 occurrences de chacun des 18 mots-cibles en parole claire, suivies de leur correction suite à une incompréhension simulée par l'expérimentatrice (réparties en 2 occurrences pour les 3 types d'erreur : sur le voisement, et sur les 2 autres lieux d'articulation).

Les données ont été étiquetées manuellement sous Praat en repérant le début et la fin des mots-cibles (t0 et t7), l'instant de disparition du 2ème formant (F2) à la fin de la voyelle précédente (t1) et celui de réapparition du F2 au début de la voyelle suivante (t6), pour les consonnes non voisées, l'instant de disparition du voisement après la fin de la voyelle précédente (t2) et celui de réapparition du voisement avec ou après le bruit de plosion (t5), enfin les instants de début (t3) et de fin (t4) du bruit (plosion+friction) émis au relâchement de l'occlusion.
A l'aide de scripts développés sous Matlab, différents descripteurs ont ensuite été extraits du signal audio sur ces différents intervalles de temps : la durée de la consonne (t7-t1), de la phase d'occlusion (t3-t1) et du bruit (t4-t3), le VOT, tel que défini par Lisker (t5-t3), le pourcentage de temps durant lequel les plis vocaux continuaient de vibrer pendant la phase d'occlusion pour les consonnes non voisées (taux de voisement : (t2-t1)/(t3-t1), l'intensité acoustique moyenne du voisement, s'il y en a, pendant la phase d'occlusion, et l'intensité du bruit.

Les analyses statistiques ont été réalisées avec le logiciel R, de façon séparée pour tester :

l-l'effet de parler plus clairement en situation de communication perturbée. Pour chaque descripteur, nous avons modélisé les données à l'aide d'un modèle linéaire incluant le facteur Condition (2 niveaux : APN et SP), le facteur Consonne (6 niveaux : /p/, /t/, /k/, /b/, /d/, /g/) et le facteur Voyelle (3 niveaux : /a/, /i/, /u/). Les mots n'étaient pas appariés entre les conditions SP et APN.

2-l'effet de corriger un mot suite à une incompréhension de l'interlocutrice. Pour tenir compte de l'appariement entre la première occurrence d'un mot et sa deuxième occurrence répétée, nous avons choisi de considérer comme variables dépendantes la variation de nos descripteurs (Δ) entre les conditions APN et APC. Pour chacune de ces variations de descripteur, nous avons modélisé les données à l'aide d'un modèle linéaire incluant le facteur Voyelle (3 niveaux : /a/, /i/, /u/) et le facteur Type d'erreur (18 niveaux : consonne voisée prise pour une non voisée, non voisée prise pour une voisée, labiale prise pour une dentale, labiale prise pour une vélaire, etc…).

Suivant la même procédure que celle développée avec des experts statisticiens, appliquée à l'analyse de plusieurs jeux de données précédents (Bourne et al. 2016), nous avons commencé par simplifier chaque modèle en excluant toute interaction non significative entre les facteurs. Une fois le modèle simplifié, nous avons vérifié sa validité en examinant ses résidus. Nous avons ensuite réalisé des tests de modèle emboités (ou Likelihood Ratio Test) pour tester la significativité de chaque facteur ou de leurs interactions. Enfin, nous avons réalisé des tests post-hoc pour examiner le contraste plus spécifique entre certaines conditions, en appliquant des corrections de Bonferroni pour les comparaisons multiples.

3 Résultats

Les analyses statistiques ont montré que la locutrice allongeait significativement la durée de ses consonnes occlusives lorsqu'elle parlait plus clairement en situation perturbée (APN vs. SP : Δ=+41ms, $p<0.0001$), avec un allongement semblable quels que soient la consonne occlusive et le contexte vocalique (modèle DuréeConsonne ~ Condition + Voyelle + Consonne). La durée des consonnes était encore plus allongée lors d'une correction, suite à une incompréhension de l'interlocutrice (APC vs. APN : Δ=+66ms, $p<0.0001$), de nouveau sans effet significatif du contexte vocalique, de la consonne ou du type d'erreur sur cet allongement (modèle ΔDuréeConsonne ~ 1). Une analyse temporelle plus détaillée a montré que cet allongement de la consonne provenait essentiellement d'un allongement significatif de la phase d'occlusion en parole claire (APN vs. SP : Δ=+44ms, $p<0.0001$ (modèle DuréeOcclusion ~ Condition + Consonne)) et lors d'une correction (APC vs. APN : Δ=+52ms en moyenne pour les consonnes en contexte /a/ et /i/, $p<0.0001$; Δ=+93ms en moyenne pour les consonnes en contexte /u/, $p<0.0001$ (modèle ΔDuréeConsonne ~ Voyelle)).

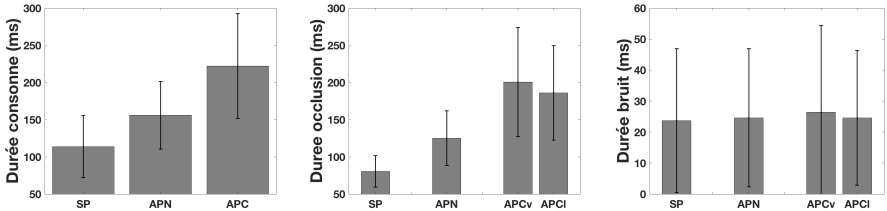

FIGURE 2: Evolution de la durée des consonnes occlusives, de leur phase d'occlusion et de leur bruit de plosion, lorsque la locutrice parlait de façon conversationnelle, dans une condition non perturbée (SP), lorsqu'elle parlait plus clairement en situation perturbée (APN), et lorsqu'elle corrigeait une erreur perceptive de son interlocutrice, commise sur le trait de voisement (APCv) ou sur le trait d'articulation (APCl).

Lors de la correction, l'allongement n'était en tout cas pas significativement affecté par le type d'erreur commise par l'auditeur (erreur sur le trait de voisement vs. sur le lieu d'articulation).
L'allongement de la consonne n'avait en revanche aucun effet significatif sur la durée du bruit de plosion, celui-ci ne variant significativement ni en parole claire, ni lors d'une correction.

Les consonnes occlusives voisées montrent une vibration des plis vocaux tout le temps de cette phase d'occlusion, tandis que les occlusives non voisées ne montrent pas de vibration, ou bien durant une courte durée après la voyelle précédente. Cette énergie basse fréquence « résiduelle » pourrait entraîner une certaine confusion perceptive, donnant à l'auditeur l'impression de voisement. C'est pourquoi nous nous attendions à ce que la durée de ce voisement résiduel et son intensité acoustique, soit diminuées lorsqu'un locuteur cherche à améliorer son intelligibilité.
En accord avec nos prédictions, notre locutrice tendait à diminuer la durée de ce voisement résiduel durant la phase d'occlusion pour ses consonnes non voisées, de façon non significative lorsqu'elle parlait plus clairement (APN vs. SP : Δ= -4.6% (modèle TauxVoisementOcclusion ~ Voyelle + Consonne)) mais de façon plus marquée lors d'une correction (APC vs. APN : Δ= -5.5%, p=0.0001), sans pour autant que cette réduction soit significativement plus marquée lorsque l'erreur commise par l'interlocutrice concernait le trait de voisement, par rapport à une erreur sur le lieu d'articulation (modèle ΔTauxVoisementOcclusion ~1).
En revanche, contrairement à nos prédictions, la locutrice augmentait globalement l'intensité moyenne du voisement pendant la phase d'occlusion lorsqu'elle parlait plus clairement, qu'il s'agisse du voisement « normal » des consonnes voisées ou du voisement résiduel des non voisées (APN vs. SP : Δ= +6.6 dB, p<0.001 (modèle IVoisementOcclusion ~ Condition + Voyelle + Consonne)). A l'inverse, lors d'une correction, la locutrice tendait plutôt à diminuer l'intensité du voisement pendant la phase d'occlusion, que ce soit pour les consonnes voisées ou non voisées (APC vs. APN : Δ= - 1.7 dB, p<0.001), avec cependant une tendance plus marquée lorsqu'il s'agissait de consonnes non voisées ayant été incorrectement perçues comme voisées par l'interlocutrice (cf. Figure 3).

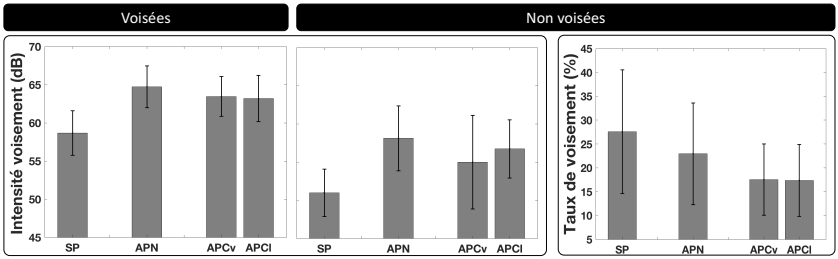

FIGURE 3: Evolution de l'intensité moyenne du voisement pendant la phase d'occlusion pour des consonnes occlusives voisées, ainsi que de la durée (relative) de voisement résiduel après la voyelle précédente durant la phase d'occlusion et de son intensité pour des consonnes occlusives non voisées, lorsque la locutrice parlait de façon conversationnelle, dans une condition non perturbée (SP), lorsqu'elle parlait plus clairement en situation perturbée (APN), et lorsqu'elle corrigeait une erreur perceptive de son interlocutrice, commise sur le trait de voisement (APCv) ou sur le trait d'articulation (APCl).

Le VOT, indice déterminant pour la perception du trait de voisement, est positif pour les consonnes occlusives du Français, plus long pour les consonnes vélaires [k] que les consonnes labiales et dentales [p t] (Lisker et Abramson, 1964). Contrairement à ce que nous aurions pu attendre, notre locutrice n'allongeait pas le VOT de toutes ses consonnes non voisées lorsqu'elle parlait plus clairement. Le VOT était significativement raccourci pour les consonnes /p/ et /t/ (APN vs. SP : Δ=- 6ms, p<0.0001) et ne variait pas significativement pour les consonnes /k/ (modèle VOT ~ Condition*Consonne + Voyelle*Consonne). Lors d'une correction (APC vs. APN), le VOT était

légèrement allongé pour les consonnes non voisées en contexte /i/ (Δ=+8ms, p=0.014) et ne montrait pas de variation significative pour les autres contextes vocaliques (modèle ΔVOT ~ Voyelle). De façon intéressante, le VOT montrait plutôt une tendance à la diminution pour les consonnes /b/ (pour lesquelles il est déjà plutôt court) et à l'augmentation pour les consonnes /k/ (pour lesquelles il est déjà plutôt long). Sa variation n'était pas significativement affectée par le type d'erreur perceptive commise par l'interlocutrice, portant sur le trait de voisement ou sur le lieu d'articulation de la consonne.

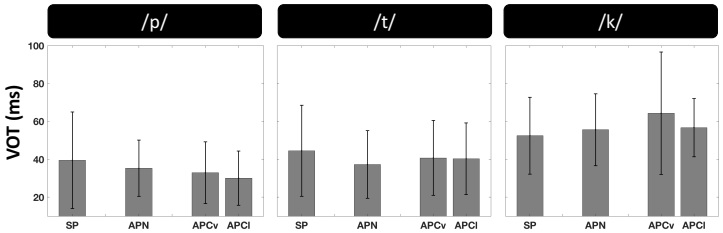

FIGURE 4: Evolution du VOT (délai d'établissement du voisement) des consonnes occlusives non voisées lorsque la locutrice parlait de façon conversationnelle, dans une condition non perturbée (SP), lorsqu'elle parlait plus clairement en situation perturbée (APN), et lorsqu'elle corrigeait une erreur perceptive de son interlocutrice, commise sur le trait de voisement (APCv) ou sur le trait d'articulation (APCl).

Les différences d'intensité et de spectre du bruit de plosion, enfin, renseignent sur le lieu d'articulation de la consonne (Stevens et Blumstein, 1978). Conformément à la littérature, nous avons effectivement bien observé chez notre locutrice des bruits de plosion assez faibles pour les consonnes occlusives labiales. En revanche, l'intensité des bruits de plosion s'est avérée relativement comparable pour ses consonnes occlusives dentales et vélaires. Dans tous les cas, notre locutrice augmentait significativement l'intensité du bruit de plosion de ses consonnes occlusives lorsqu'elle parlait plus clairement (APN vs. SP : Δ=+7.9 dB, p<0.0001), de façon comparable pour toutes les consonnes (modèle IBruitPlosion ~ Condition + Voyelle*Segment). Lors d'une correction (APC vs. APN), l'intensité du bruit de plosion était encore davantage renforcée pour des consonnes occlusives en contexte vocalique /u/ (+3.4 dB, p=0.005) mais ne variait pas significativement pour les autres contextes (modèle ΔIBruitPlosion ~ Voyelle). Les variations d'intensité du bruit de plosion ne dépendaient pas significativement de l'erreur perceptive commise par l'interlocutrice, en particulier du lieu d'articulation incorrectement perçu (labial, dental ou palatal).

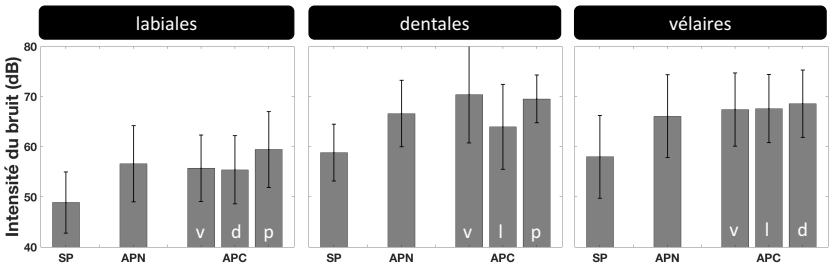

FIGURE 5: Evolution de l'intensité du bruit de plosion des consonnes occlusives labiales [p b], dentales [t d] ou vélaire [k g] lorsque la locutrice parlait de façon conversationnelle, dans une condition non perturbée (SP), lorsqu'elle parlait plus clairement en situation perturbée (APN), et lorsqu'elle corrigeait une erreur perceptive de son interlocutrice, commise sur le trait de voisement (APCv) ou sur les deux autres lieux d'articulation que celui de la consonne produite (APCl pour un lieu d'articulation incorrectement perçu comme labial, APCd pour dental, APCp pour palatal).

4 Discussion et conclusion

Sur quels indices acoustiques nous basons-nous en tant que locuteur ou auditeur, pour distinguer les différentes catégories de consonnes occlusives, et lesquels d'entre eux renforçons-nous lorsque nous devons améliorer notre intelligibilité en situation de communication perturbée ou lorsque notre interlocuteur nous a mal compris ?
Avec toutes les réserves liées au fait que nous avons pour l'instant exploré cette question sur une seule locutrice et sur un ensemble restreint de descripteurs acoustiques, nous pouvons néanmoins déjà dire que pour améliorer son intelligibilité, la locutrice de notre expérience pilote a montré plusieurs modifications significatives de la production de ses consonnes occlusives, allant dans le même sens en parole claire (APN vs. SP) et lors d'une correction (APC vs. APN), avec une modification plus marquée lors d'une correction: un allongement de la durée de ses consonnes occlusives provenant essentiellement d'un allongement de leur phase d'occlusion, une diminution de la durée du voisement résiduel à la fin de la voyelle précédente pendant la phase d'occlusion de ses consonnes non voisées, une diminution du VOT sur les consonnes /p/ et une augmentation sur les consonnes /k/, une augmentation de l'intensité du bruit de plosion-friction au relâchement de l'occlusive. Ces modifications acoustiques sont globalement conformes à nos attentes et peuvent s'interpréter en termes de stratégie communicationnelle visant à améliorer l'audibilité des indices et leur temps de récupération (intensité du bruit, durée de la phase d'occlusion), mais visant également possiblement à renforcer le contraste inter-catégoriel (VOT des consonnes non voisées labiales et vélaires, durée du voisement résiduel pendant la phase d'occlusion des consonnes non voisées).

Cette distinction des différentes catégories de consonnes occlusives est-elle particulièrement claire lorsque ces indices sont « prototypés », prenant des valeurs très spécifiques, ou bien lorsque ces indices sont « contrastés », prenant des valeurs les plus distinctes possibles entre deux catégories phonologiques ?
Contrairement à nos attentes, seul un des descripteurs examinés, la durée de voisement résiduel à la fin de la voyelle précédente pendant la phase d'occlusion des consonnes non voisées, a été davantage diminué par la locutrice suite à des erreurs perceptives de son interlocutrice commises sur le trait de voisement, par rapport à des erreurs commises sur le lieu d'articulation. La modification des autres descripteurs ne s'est pas montré affectée par le type d'erreur perceptive commise par l'interlocutrice, laissant penser que lors d'une correction, notre locutrice ne cherchait pas tant que cela à renforcer des contrastes acoustiques.

La prochaine étape de ce projet consistera, très naturellement, à étendre ces premières analyses à un ensemble plus complet de descripteurs acoustiques et articulatoires des consonnes occlusives (spectre du bruit, transitions formantiques, degré de compression des lèvres lors d'une occlusion labiale, etc) et à généraliser (ou non) ces observations à une cohorte de 10-15 locuteurs.

Remerciements

Nous remercions Christophe Savariaux et Frédéric Elisei pour leur aide lors de la mise au point de l'expérience et de l'acquisition des données.
Cette recherche est financée par l'Agence Nationale de la Recherche (Projet StopNCo : Effort et coordination dans la production des consonnes occlusives ; ANR-14-CE30-0017; Maëva Garnier).

Références

ABDELLI-BERUH. N. (2004). The Stop Voicing Contrast in French Sentences: Contextual Sensitivity of Vowel Duration, Closure Duration, Voice Onset Time, Stop Release and Closure Voicing, Journal: Phonetica, vol. 61, no. 4, pp. 201-219.

BOURNE, T., GARNIER M., SAMSON A. (2016). Physiological and acoustic characteristics of the male music theatre voice. The Journal of the Acoustical Society of America, 140(1), 610-621

CARLSON, R., GRANSTRÖM, B., PAULI, S. (1972). Perceptive evaluation of segmental cues. STL/QPSR 1/1972, Stockholm, Roy. Inst.Technol.; 18-24.

CHO, T., LADEFOGED, P. (1999). Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics, 27(2), 207-229.

CRYSTAL T., HOUSE A. (1988). A note on the variability of timing control. Journal of Speech and Hearing Research, 31, 497-502.

DELATTRE P., LIBERMAN A., COOPER F. (1955). Formant transitions and loci as acoustic correlates of place of articulation in American fricative consonants. Studia Linguistica, 16, 104-121.

DELATTRE, P. (1962) Some factors of vowel duration and their cross-linguistic validity, JASA, 34, 1141-1142.

DORMAN M. F., STUDDERT-KENNEDY M., RAPHAEL L. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. Perception and Psychophysics, 22, 109-122

FORREST, K., WEISMER, G., MILENKOVIC, P., DOUGALL, R. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. The Journal of the Acoustical Society of America, 84(1), 115-123.Stevens and Blumstein 1978

FRANCIS, A., KAGANOVICH, N., DRISCOLL-HUBER, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. The Journal of the Acoustical Society of America, 124(2), 1234-1251.

GRAVEL, J., OHDE, R. (1983). Perception of stop place of articulation: Effects of stimulus amplitude. American Speech-Language-Hearing Association, 25, 101.

HARRIS, K.S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech 1, 1-7.

LIBERMAN A., DELATTRE P., COOPER F., GERSTMAN L. J., (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 68, 1-13.

LISKER, L. (1957). Closure duration and the intervocalic voiced- voiceless distinction in English. Language 33, 42-49.

LISKER, L., ABRAMSON, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20(3), 384-422.

LISKER, L., ABRAMSON,A. (1970). The voicing dimension: some experiments in comparative phonetics. Proc. of the 6th Int. Cong. of Phonetic Sciences, Prague 1967; Prague: Academia, 1970; 563-567.

LISKER, L. (1975). Is it VOT or a first formant transition detector? Acoust.Soc.Am. 57, 1547-1551

LOUSADA, M., JESUS, L., PAPE, D. (2012). Estimation of stops' spectral place cues using multitaper techniques. DELTA 28(1), 1-26.

MILLER, J. (1981). Phonetic perception: Evidence for context- dependent and context-independent processing. J.Acoust.Soc.Am. 69, 822-831.

OHALA, J., RIORDAN,C. (1979). Passive vocal tract enlargement during voiced stops" in Speech Comunication Papers, J. Wolf and D. Klatt eds.; Acoust.Soc.Am.: New York; 89-92.

OHDE, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. The Journal of the Acoustical Society of America, 75(1), 224-230.

POLS, L., SCHOUTEN, M. (1985). Plosive consonant identification in ambiguous sentences. J.Acoust.Soc.Am. 78, 33-39.

REPP, B., LIBERMAN, A., ECCARDT, T., PESETSKY, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 4(4), 621.

SERNICLAES, W. (1984). Fenêtre de prélèvement temporel des indices d'occlusives. Dans Actes des XXIèmes Journées d'Etudes sur la Parole, 67-78.

STEVENS, K., KLATT, D. (1974). Current models of sound sources for speech. In Ventilatory and phonatory control systems: and international symposium. Oxford University Press, New York.

STEVENS, K., MANUEL, S., MATTHIES, M. (1999). Revisiting place of articulation measures for stop consonants: Implications for models of consonant production. In Proceedings of the International Congress of Phonetic Sciences (pp. 1117-1120).

STUDDERT-KENNEDY M. (1990). Language development from an evolutionary perspective. Haskins Laboratories Status Report on Speech Research, 101-102, 14-27.

SUMMERFIELD, Q., HAGGARD, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. The Journal of the Acoustical Society of America, 62(2), 435-448.

WILLIAMS, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. Psychological Bulletin, 84(5), 963.

WINN, M., CHATTERJEE, M., IDSARDI, W. (2013). Roles of Voice Onset Time and F0 in Stop Consonant Voicing Perception: Effects of Masking Noise and Low-Pass Filtering. Journal of Speech, Language, and Hearing Research, 56(4), 1097-1107.