

Université de Montréal

Analyse acoustique de la conviction exprimée dans des réponses de sondage : Un raffinement  
des enquêtes téléphoniques

par

Anne-Marie Lespérance

Département de linguistique et de traduction

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de Maîtrise ès arts

en Linguistique

Juin 2014

© Anne-Marie Lespérance, 2014

Université de Montréal

Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Analyse acoustique de la conviction exprimée dans des réponses de sondage : Un raffinement  
des enquêtes téléphoniques

Présentée par :

Anne-Marie Lespérance

a été évaluée par un jury composé des personnes suivantes :

Patrick Drouin, président-rapporteur

Victor Boucher, directeur de recherche

Patrick Fournier, membre du jury

André Clas, membre du jury

## Résumé et mots clés

Les sondages téléphoniques visent habituellement l'enregistrement de réponses selon leur contenu linguistique littéral (« oui », « jamais », nom de politiciens, etc.). Cependant, les réponses orales contiennent également des aspects acoustiques qui expriment le degré de conviction des répondants. Dans les techniques actuelles de sondage, cette information, dite « extralinguistique », n'est pas retenue bien qu'elle puisse servir à la prédiction des comportements des répondants. Le présent mémoire a pour objectif de démontrer que des mesures acoustiques de la conviction exprimée par les répondants peuvent servir à améliorer la prédiction de comportements telle qu'estimée par les techniques de sondage. Cette démonstration se fait en deux étapes et implique deux expériences. L'Expérience 1, réalisée en laboratoire ( $n = 25$ ), avait pour but d'isoler les variables vocales associées à l'expression de la conviction. Ce test a permis d'isoler des variables mesurables pertinentes et les distributions catégoriques de valeurs permettant une catégorisation des réponses selon un « degré de conviction ». L'Expérience 2 a porté sur l'application de ces mesures acoustiques dans l'analyse de réponses enregistrées lors d'un sondage téléphonique préélectoral fait par la firme *Léger* en 2010 (un corpus de 400 entrevues). Cette deuxième expérience a démontré qu'une catégorisation des réponses nominatives ( $n = 188$ ) sur la base des variables associées à l'expression de la conviction (telles que définies à l'Expérience 1) peut améliorer la prédiction des comportements des électeurs. Les résultats suggèrent que les réponses orales à un sondage contiennent des indices acoustiques mesurables qui expriment un degré de conviction relié à une volonté de produire certains comportements. La nature de ces variables présente aussi la possibilité d'une détection automatique.

Mots clés : Sondages politiques, prédiction, conviction, mesures acoustiques

## **Abstract and keywords**

Surveys by phone usually involve the recording of responses in terms of their literal linguistic content ("yes", "never", politician name, etc.). However, oral responses also possess acoustic aspects that express a degree of conviction of respondents. In current survey techniques, this "extralinguistic" information is not considered although it can be used to predict the respondent's behavior. This paper aims to demonstrate that acoustic measures of the conviction expressed by respondents can be used to improve the prediction of behaviors as estimated by survey techniques. This demonstration is done in two steps involving two experiments. Experiment 1, conducted in a laboratory setting ( $n = 25$ ), was designed to isolate the vocal variables associated with the expression of the conviction. This first test served to isolate the relevant measurable variables and "categorical" values that could be used to categorize oral responses. Experiment 2 involved the application of these acoustic measures in the analysis of responses recorded during a pre-election telephone survey conducted by *Leger* in 2010 (a corpus of 400 interviews). This second experiment showed that a categorization of oral responses ( $n = 188$ ) based on the basis of values of expressed conviction (as defined in Experiment 1) can improve the prediction of the behavior of voters. The results suggest that oral-survey responses contain measurable acoustic cues that express a degree of conviction, which can be associated with a willingness to produce certain behaviors. The nature of the variables also presents the possibility of an automatic detection. The nature of the variables also presents the possibility of an automatic detection.

Keywords: Political survey, prediction, conviction, acoustic measurements

# Table des matières

Liste des tableaux.....	i
Liste des figures .....	ii
Liste des abréviations.....	iii
Remerciements.....	iv
1. Introduction.....	1
1.1 Pertinence d'une étude sur la catégorisation de l'expressivité en sondage.....	1
2. Les indices de la conviction dans les réponses orales .....	4
2.1 Le lien entre l'affect et l'émotion : pour clarifier les termes.....	5
2.2 Sur les réactions associées à la conviction et leurs effets acoustiques dans la parole.....	6
2.3 Les variables de l'expression de la conviction dans le contexte d'un sondage.....	9
2.4 Le problème général de la catégorisation de l'expressivité : une approche objective basée sur les observations <i>a posteriori</i> .....	10
3. Objectifs spécifiques et Hypothèses .....	11
3.1 Expérience 1 .....	11
3.2 Expérience 2 .....	11
4. Expérience 1 .....	12
4.1 Méthode.....	12
4.1.1 Participants.....	12
4.1.2 Stimuli.....	13
4.1.3 Procédures.....	14
4.1.4 Analyse des réponses : mesures acoustiques .....	15
4.1.5 L'analyse statistique.....	20
4.2 Résultats .....	21

4.3	Discussion .....	28
5.	Expérience 2 .....	30
5.1	Méthode.....	30
5.1.1	Participants.....	30
5.1.2	Stimuli.....	31
5.1.3	Procédures et enregistrements audio.....	31
5.1.4	Analyse des réponses à la question cible .....	32
5.2	Résultats .....	34
5.2.1	Les variables servant au classement .....	36
5.2.2	Les valeurs de classement.....	39
5.2.3	La validation des variables de classement .....	42
5.3	Discussion .....	44
6.	Discussion générale et conclusion .....	46
7.	Bibliographie .....	49
	Annexe 1 : documents pour l'Expérience 1 .....	53
	Annexe 2 : documents pour l'Expérience 2 .....	58

## Liste des tableaux

Tableau I. *Résumé des principaux corrélats de l'expression d'émotions dans la parole :*

*tableau de Murray et Arnott, 1993, tiré de l'article de Gustafson-Capková, 2001.*

Tableau II. *Tests t pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc A.*

Tableau III. *Tests t pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc B.*

Tableau IV. *Tests t pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc C.*

Tableau V. *Tests t pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc D.*

Tableau VI. *Tests t pour échantillons appariés des variables acoustiques en contexte expressif et neutre.*

Tableau VII. *Tableau comparatif des résultats bruts des sondages et des analyses acoustiques sur les élections municipales à Winnipeg, 2010.*

## Liste des figures

Figure 1. *Exemple d'une segmentation pour le lexème « non ».*

Figure 2. *Exemple d'analyse acoustique de la fréquence fondamentale pour la réponse « non ».*

Figure 3. *Exemple d'analyse acoustique de la fréquence fondamentale contenant des valeurs erronées dans la réponse « jamais ».*

Figure 4. *Exemple d'analyse acoustique de l'amplitude pour la réponse « non ».*

Figure 5. *Exemple d'analyse du temps de réaction de la réponse « non ».*

Figure 6. *Exemple de segmentation de la voyelle [a] dans « Sam ».*

Figure 7. *Histogrammes illustrant la distribution des valeurs différenciées observées pour l'intonation et l'amplitude et les valeurs de coupe estimées servant à catégoriser un répondant selon sa conviction.*

Figure 8. *Histogrammes illustrant la distribution des valeurs brutes observées pour la durée et le temps de réaction et les valeurs de coupe estimées servant à catégoriser un répondant selon sa conviction.*



## Liste des abréviations

dB : décibel

$F_0$  : fréquence fondamentale

Hz : hertz

max. : maximum

min. : minimum

ms : milliseconde

$n$  : nombre de sujets

$p$  : probabilité d'erreur

sec. : seconde

syll. : syllabe

% : pour cent

$\bar{x}$  : moyenne

*é.t.* : écart type

## **Remerciements**

Merci à toute l'équipe du Laboratoire de sciences phonétiques pour l'aide apportée à la réalisation de ce projet. Un merci tout spécial à ma famille et à mes amis pour leur patience, leur appui et leurs encouragements.

# 1. Introduction

## 1.1 Pertinence d'une étude sur la catégorisation de l'expressivité en sondage.

Un sondage d'opinion se base régulièrement sur la collecte par téléphone de données du type « oui », « non », « ne sait pas », en réponse à un questionnaire oral. Or, ces réponses orales contiennent davantage d'information que leur sens littéral. L'aspect acoustique des réponses véhicule aussi des informations dites « extralinguistiques » sur l'affectivité pouvant refléter la conviction ou l'hésitation du répondant. Il s'agit en fait d'aspects identifiables à partir d'indices prosodiques dans la voix. Toutefois, ces informations associées à l'expressivité ne sont habituellement pas recueillies par les sondeurs. En fait, dans un sondage téléphonique, il est rarement demandé à un répondant de faire état de ses réactions comme, par exemple, son degré de conviction. Ces informations sont perdues bien qu'elles puissent servir à la prédiction des comportements, ce qui constitue l'un des objectifs du sondage. L'objectif d'un sondage électoral est de prédire la distribution des votes selon les candidats ou les partis politiques en lice et, dans ce sens, il est pertinent de considérer toute information pouvant améliorer la prédiction des comportements des électeurs. L'efficacité prédictive de l'information recueillie dans un sondage se vérifie indirectement par l'observation des comportements de la population en réponse au sondage, comme la distribution des votes le jour des élections (Guilbert, Haziza, Ruiz-Gazen, Tillé, 2008). Cette procédure constitue en fait une vérification empirique *a posteriori* de l'outil de prédiction. Par exemple, si une majorité des répondants ont indiqué leur appui à un parti, mais que celui-ci n'est pas élu, l'efficacité des questions pourrait être remise en doute. En somme, au-delà de l'information littérale ou linguistique contenue dans une réponse à un sondage téléphonique, il y aurait vraisemblablement un avantage à mesurer l'information extralinguistique exprimant la volonté du répondant à produire un comportement qui fait l'objet de l'enquête. Aussi, dans plusieurs types de sondage, dont le sondage électoral, il serait tout à fait possible de vérifier *a posteriori* l'efficacité de mesures sur l'information extralinguistique dans la prédiction des comportements d'une population.

Toutefois, bien qu'on puisse entrevoir les avantages évidents de certaines mesures de l'expressivité vocale dans la prédiction de comportements, les techniques de sondage actuelles

accordent peu d'attention à cette information extralinguistique contenue dans la voix des répondants. Le présent travail prend pour objet cette information extralinguistique contenue dans les réponses à un sondage téléphonique. Nous pensons que l'omission de cette information dans les techniques actuelles relève principalement du manque d'outils d'analyse acoustique dans les firmes de sondage pour quantifier des données orales. Il y a aussi une absence d'études sur les variables acoustiques. Le problème, en fait, ne porte pas sur la désignation de variables associées à l'expressivité. Comme nous le verrons, il existe plusieurs travaux sur les aspects vocaux liés à l'affect (le terme sera défini dans la section 2.1). Cependant, il y a un problème théorique central de catégorisation de l'expressivité, en particulier pour l'expression d'un « degré de conviction ». En considérant ce problème de catégorisation, il est clair que l'on peut difficilement référer aux jugements perceptifs (des répondants ou du sondeur). Toute évaluation subjective faite par un répondant ou un évaluateur entraînerait inévitablement un manque de fiabilité inter-juge ce qui nuirait à l'efficacité prédictive de la mesure. Il importe donc non seulement d'*établir* les variables expressives qui permettent d'améliorer la prédiction des comportements, mais aussi de trouver un principe de mesure et de *catégorisation objective* des réponses.

Dans cette optique, on propose un travail en deux étapes qui a pour **objectif général** de démontrer que des mesures acoustiques de la conviction exprimée dans les réponses à un sondage téléphonique peuvent servir à améliorer la prédiction de comportements tels qu'estimés par les techniques de sondage. Cette démonstration se fait essentiellement en deux étapes. L'**objectif spécifique de la première étape** est d'isoler, au moyen d'une expérience contrôlée (Expérience 1), les variables vocales associées à l'expression de la conviction. Cette expérience permettra de dégager des variables acoustiques que l'on peut mesurer et qui présentent des distributions « catégoriques » de façon à répondre au problème de catégorisation objective des réponses. Ayant défini les variables acoustiques pertinentes, l'**objectif spécifique de la deuxième étape** est de démontrer que ces variables acoustiques permettent de catégoriser autrement les réponses téléphoniques d'un sondage électoral fait par Léger (anciennement *Léger Marketing*) en 2010 et ainsi d'améliorer la prédiction de résultats électoraux issus de ce sondage (ce qui constitue, en fait, l'Expérience 2).

Dans le travail qui suit, on précise les hypothèses associées aux objectifs ci-dessus et leur cadre théorique en six sections. À la section suivante (**section 2**), on présente une recension de la littérature sur les indices associés à l'expression vocale de la conviction. On propose une clarification des termes que sont l'« affect », l'« émotion » et l'« expression de la conviction ». On aborde aussi les différentes approches pour catégoriser et explorer les différentes variables de l'expression de la conviction. Dans les **sections 3, 4 et 5**, on expose les hypothèses et les motivations de deux expériences au centre de notre démonstration. L'Expérience 1 répond au problème de définition opérationnelle d'une mesure vocale catégorique. L'Expérience 2 présente une démonstration de l'applicabilité des variables pour la prédiction de comportements lors d'un sondage. Enfin, la **section 6** combine une discussion générale et une conclusion sur les variables et ce qu'elles permettent de faire, soit d'arriver à augmenter la capacité prédictive des sondages téléphoniques.

## **2. Les indices de la conviction dans les réponses orales**

Les réponses orales à un sondage comportent toujours des informations linguistiques que l'on peut transcrire, par exemple « oui/non ». Elles comportent aussi des informations non linguistiques, que l'on transcrit difficilement et qui reflètent des attributs du locuteur et de la situation de communication. Dans leur récente étude sur la perception prosodique, Nygaard, Herold, et Namy (2009) ont constaté que l'« intention communicative » d'un locuteur n'est pas seulement de transmettre des informations linguistiques. En fait, différents indices prosodiques véhiculent des informations relatives non seulement à l'âge, au sexe ou à la condition physique du locuteur, mais aussi à l'expression de l'émotion et du degré de conviction de l'individu (voir Mozziconacci, 2002). Comme le remarquent certains auteurs, ces indices associés à l'expressivité émotionnelle présentent un aspect « iconique » (Nygaard et coll., 2009) en ce qu'ils reflètent des réactions ou des réponses primaires. En fait, il est bien connu que l'expression d'émotions est liée aux systèmes limbiques et autonomes (Izard, 1972; Ledoux, 1996, on y reviendra sous peu). Les variations d'indices acoustiques concernant l'émotivité ne sont donc pas « symboliques » au sens saussurien du terme, mais bien iconiques, comme le souligne Nygaard et coll. (2009), en ce que ces indices reflètent fondamentalement des réponses ou des réactions d'ordre physiologiques. Par exemple, le fait de produire de la parole à forte intensité ou avec une fréquence fondamentale ( $F_0$ ) élevée pour exprimer la surprise ou la peur n'est pas fortuit, mais universel chez les humains. Cependant, il demeure très difficile d'interpréter et de catégoriser les indices d'émotions exprimées, surtout lorsqu'il s'agit d'aspects subtils comme la conviction. Les prochaines sections présentent un survol de ces problèmes d'analyse de l'émotivité dans la parole et des questions entourant l'expression de la conviction, dans le domaine du sondage. Soulignons que l'application d'indices vocaux dans ce domaine n'a pas fait l'objet de recherche jusqu'à maintenant.

## 2.1 Le lien entre l'affect et l'émotion : pour clarifier les termes

On définit l'émotion en tant qu'un ensemble de réponses affectives et physiologiques transitoires, d'intensité variable, associées à un événement dans l'environnement sensoriel d'un individu (Larousse, 2014). Ces réponses physiologiques ne se limitent pas à des réactions systémiques comme, par exemple, l'augmentation du rythme cardiaque ou la sudation. Elles s'appliquent également aux mouvements et aux vocalisations. Par rapport au terme « émotion », notons que la notion d'« affect », ou de réponses affectives, est plus englobante. Pour Juslin et Scherer (2005), l'affect réfère aux émotions, au stress, à l'humeur et aux traits affectifs de la personnalité qui influencent la motivation et le comportement. Toutes ces manifestations ont un aspect « affectif » qui se distingue d'un état passif ou neutre. On remarque cependant que, dans la littérature, il n'y a pas de définition acceptée sur ce que sont les émotions. Il n'y a pas non plus de consensus sur les différents types d'émotions ou leur nombre. En fait, les études sur l'expression vocale des émotions font typiquement usage de trois approches dans la catégorisation des émotions.

La première approche, qui est sans doute la plus répandue dans la littérature, s'axe sur le présupposé qu'il y a des catégories discrètes d'émotions universelles (voir p. ex., Izard, 1977; Plutchik, 1980; Ekman, 1982; Murray et Arnott, 1993). En général dans cette approche, on considère qu'il y a cinq ou six émotions dites « primaires » (surtout la joie, la colère, la peur, le dégoût, la tristesse, la surprise) et des combinaisons de celles-ci. Selon Izard (1972), chaque émotion est unique. Cette catégorisation se base principalement sur l'observation des expressions faciales.

La deuxième approche est « dimensionnelle » et a été proposée par différents auteurs dont les principaux sont Nygaard et coll. (2009); Cowie et Cornelius (2003); Cohen, Hong et Guevara (2009); Bachorowski (1999); Schröder (2001) et Russell, Bachorowski et Fernandez-Dols (2003). Dans cette perspective, on caractérise les émotions selon un degré d'intensité et d'activation dans un continuum correspondant à des réactions physiologiques liées au rythme cardiaque, à la transpiration, etc. Ainsi, les émotions se trouvent à être définies par rapport à différents axes (positif, négatif, neutre; actif, passif) reflétant deux ou trois dimensions (Cowie et coll., 2003).

La troisième tendance réfère à une approche constructiviste où l'on conçoit les émotions comme des réactions acquises, propres à chaque culture (Pantic, Sebe, Cohn, et Huang, 2005 ; Mozziconacci, 2002). Selon Pantic et coll. (2005), l'étiquette donnée à une émotion dépend de la culture et les émotions varient d'une culture à l'autre et d'une personne à l'autre. Dans cette optique, il y a une interprétation sociale ou culturelle des indices comportementaux pour les états affectifs de sorte que les « émotions » ne réfèrent aucunement à des universaux. Cette troisième approche s'inspire des deux premières perspectives ci-dessus en ajoutant une considération de l'aspect relatif, culturel et contextuel de toute catégorisation des réponses émotives.

En dépit de ces questions de catégorisation, il est important de reconnaître que les réactions *vocales* à un quelconque événement sont des manifestations de l'affect et des émotions d'un individu (Larousse, 2014). Ainsi, ce que Vaissière (2004, 2006) appelle la « fonction expressive de la parole » réfère aux variables vocales des émotions et des affects suscités par des événements – incluant des questions de sondage. La question critique pour notre étude est de savoir si l'expression de la conviction présente des variables mesurables suffisamment catégoriques (ou spécifiques sur un continuum) afin de distinguer objectivement de brèves réponses orales dans le cadre d'un sondage.

## **2.2 Sur les réactions associées à la conviction et leurs effets acoustiques dans la parole**

Il est clair que l'expression ou le manque de la conviction apporte des informations utiles dans le cadre des méthodologies de sondage où la prévision des comportements constitue l'objectif essentiel. Les variables vocales seront décrites plus loin. Notons pour l'instant que l'aspect catégorique de l'expression d'émotions apparaît d'ores et déjà possible du moins pour certaines émotions et pour des cas assez subtils. Par exemple, certains indices acoustiques iconiques ont été proposés pour identifier la colère et la confiance (Ohala, 1994) ou des catégories plus fines telles que le sarcasme (Izard, 1972; Rockwell, 2000). Ces associations entre des mesures acoustiques et la perception de catégories d'émotions, même si celles-ci sont spécifiques à un contexte ou à une culture, pourraient s'étendre à l'intention ou la conviction. Afin de saisir le lien avec certaines variables vocales, il faut considérer le



fonctionnement du système limbique et autonome qui, comme on l'a suggéré plutôt, sous-tend certaines réactions physiologiques pouvant avoir des effets vocaux.

Les recherches d'Izard (1972) et de Ledoux (1996) ont lié l'expression d'émotions au système limbique et à l'hypothalamus. De façon générale, le système limbique assigne des valeurs positives ou négatives aux sensations. Il agit sur l'hypothalamus qui intervient dans la régulation du système nerveux autonome de façon à moduler les réactions systémiques primaires, par exemple l'assèchement de la bouche, l'accélération du rythme cardiaque, la transpiration et le rougissement. Ces réactions primaires s'observent selon différentes techniques de mesure physiologique et sont liées à l'expression d'émotions ou d'affects. Par exemple, il est connu que des observations sur le souffle, sur le pouls, ou sur la sudation servent à évaluer le stress et même l'hésitation (Greco et Roger, 2003; Monat, Averill, et Lazarus, 1972) et les émotions de base, dont la peur, la colère, la tristesse et la joie (Williams et Stevens, 1972 ; Levenson, Ekman, et Friesen, 1990; Rainville, Bechara, Naqvi, et Damasio, 2006; Balconi, Brambilla, et Falbo, 2009). Selon Leitman et coll. (2010), le cortex temporal supérieur, les amygdales, le gyrus frontal inférieur et le gyrus cingulaire sont impliqués dans l'assignation ou la perception d'émotions, car ces structures sont activées lors de l'écoute de la prosodie affective. Il n'y a toutefois pas d'études corrélatives entre des mesures physiologiques de réactions et l'expression d'émotions, ce qui serait d'une utilité centrale dans la définition opérationnelle de catégories d'émotions selon des variables vocales mesurables. En fait, selon Russell et coll. (2003), la plupart des gens peuvent inférer l'état psychologique d'un locuteur selon des changements vocaux. Aussi, certaines études permettent de dégager des variables qui s'associent à l'expression de la conviction.

Plusieurs études ont proposé des variables prosodiques acoustiques servant à caractériser l'expression des émotions (voir Léon, 1971 (pour le français); Scherer, 1986; Beller, 2007). Par exemple, Coutinho et Dikken (2013) ont démontré que la perception de la prosodie émotionnelle peut être prédite à partir d'un ensemble d'indices psychoacoustiques, dont l'amplitude, le débit, le contour de l'intonation (relatif à la fréquence fondamentale ( $F_0$ )), la netteté (*sharpness*) et l'inégalité (*roughness*). Pour les auteurs Murray et Arnott (1993), les variables les plus importantes sont la  $F_0$  (la moyenne et la variance); ensuite la durée;

l'amplitude et la qualité de la voix, telle que rapportée par Gustafson-Capková (2001). Aussi, Schröder, Cowie, Douglas-Cowie, Westerdijk, et Gielen (2001) utilisent les mêmes paramètres, mais y considèrent également l'articulation. Suivant ces travaux sur la prosodie, on retrouve souvent des grilles descriptives (Tableau I) où l'on réfère à certaines catégories usuelles d'émotions.

Tableau I. *Résumé des principaux corrélats de l'expression d'émotions dans la parole : tableau de Murray et Arnott, 1993, tiré de l'article de Gustafson-Capková, 2001.*

Variables	Emotions			
	<i>Anger</i>	<i>Happiness</i>	<i>Sadness</i>	<i>Fear</i>
Speech rate	Slightly faster	Faster or slower	Slightly slower	Much faster
Pitch average	Very much higher	much higher	Slightly lower	Very much higher
Pitch range	Much wider	Much wider	Slightly narrower	Much wider
Intensity	Higher	Higher	Lower	Normal
Voice quality	Breathy, chest tone	Breathy, blaring	Resonant	Irregular voicing
Pitch changes	Abrupt, on stressed syllables	Smooth, upward inflections	Downward inflections	Normal
Articulation	Tense	Normal	Sturring	Precise

Quant à la question centrale de l'expression de la conviction, il est intéressant de noter que Schröder et coll. (2001) ont suggéré une catégorisation des émotions impliquant l'« activation » selon que le locuteur est amené à agir ou qu'il hésite et reste passif. Cette notion d'activation, dans le contexte d'une réponse à un sondage, s'apparente à rappeler la conviction au sens que, plus un individu est convaincu dans son choix pour un candidat à une élection, plus il est prêt à passer à l'action et à voter pour l'individu en question. Un degré d'activation moindre ou une « inaction » suggèrerait, à l'inverse, un doute, une incertitude, une hésitation à passer à l'action. Dans cet ordre d'idée, Brennan et Williams (1995) notent que les hésitations pleines (angl. *filled hesitations*) accompagnées d'une montée d'intonation

dénotent souvent une incertitude de la réponse à une question, mais une baisse d'intonation permet de signaler une connaissance (Brennan et coll., 1995). Cependant, Schober et Bloom (2004) affirment que le manque de compréhension d'une question en entrevue se transmet souvent par une combinaison d'indices, dont des hésitations pleines (telles que "um", "uh"), de longues pauses, la répétition de l'élément ambigu ou incompris, et/ou la reprise hésitante de la réponse (p. ex. "y-y-yes"). De façon plus pertinente, dans une étude basée sur l'évaluation perceptive des réponses orales, Scherer, London et Wolf (1973) rapportent que le niveau de confiance des locuteurs (ce qui rejoint conceptuellement la conviction) se caractérise par une montée d'intensité, une haute  $F_0$ , une parole plus rapide et des pauses relativement plus courtes. Des résultats similaires ont été rapportés par Kimble et Seidel (1991), qui ont analysé les caractéristiques vocales. Pour Ohala (1994), une haute  $F_0$  qui chute rapidement symbolise la confiance. Enfin, Scherer et coll. (1973) et Kimble et coll. (1991) ont associé la conviction à une augmentation de l'amplitude.

### **2.3 Les variables de l'expression de la conviction dans le contexte d'un sondage**

Un nombre d'études a fait état de variables associées à l'expression de la conviction dans les réponses orales à un sondage. Deux principales variables ont été suggérées dont la durée et le temps de réponse. Ainsi, certains travaux ont démontré que, dans le cadre d'un sondage, l'hésitation avec des ajouts comme *well...*, *uh*, *um*, *maybe*, *I dunno*, *I guess*, *I think*, des silences (Schaeffer, Maynard et Cradock, 1993) ou la production de pauses (Schober et coll., 2004) dénotaient un manque de conviction. D'un autre côté, des études ont démontré que la conviction s'exprimait par la rapidité de la durée des réponses ou la brièveté du temps de réponse (Schober et coll, 2004; Schaeffer et coll, 1993). Certains auteurs ajoutent que des montées d'intensité peuvent aussi signaler la conviction (Scherer et coll., 1973; Kimble et coll., 1991). Toutefois, des données sur l'intonation et l'amplitude ne sont pas habituellement recueillies lors d'un sondage d'opinion même si, comme on l'a vu, plusieurs études suggèrent que ces variables permettent de signaler la conviction.

## **2.4 Le problème général de la catégorisation de l'expressivité : une approche objective basée sur les observations *a posteriori***

Ayant précisé les variables acoustiques possibles de l'expression de la conviction, toute étude de l'applicabilité d'une mesure acoustique fera face au problème de définir les valeurs propres à un degré de conviction pouvant prédire les comportements ou les agissements du locuteur. Par exemple, à supposer que l'on détermine que le temps de réponse et/ou une montée relative d'intensité reflètent la conviction d'un électeur face à un candidat. Il faut se demander si le degré de conviction permet de refléter la volonté d'agir des électeurs et, en fin de compte, mener à de meilleures prédictions des résultats électoraux. Une réponse à ce type de problème peut s'établir *a posteriori*, sans présupposé de catégorie « plus/moins » convaincu. On suit, en cela, la perspective de Caelen-Haumont (2004) qui suggère que la catégorisation de variables expressives s'établit objectivement *a posteriori* en observant les comportements. Cette approche est d'autant plus souhaitable que plusieurs variables associées à l'expression d'émotions se combinent (Cowie et Cornelius, 2003). En prenant cette optique dans la présente étude, on adopte une procédure qui consiste à définir d'abord, dans une première expérience visant une analyse de réponses orales, les variables acoustiques opérationnelles qui reflètent l'expression de la conviction. Ayant trouvé les variables pertinentes, on définit les valeurs qui permettent d'améliorer la prédiction des résultats d'un sondage fait par la firme de recherche *Léger* (étude réalisée en 2010) par rapport aux résultats réels obtenus dans un contexte électoral. Il s'agit, en somme d'une vérification *a posteriori* dans le cadre d'une seconde expérience constituée par des données déjà acquises. Les paragraphes suivants clarifient les objectifs spécifiques de chaque expérience en question.

### **3. Objectifs spécifiques et Hypothèses**

#### **3.1 Expérience 1**

À partir de l'idée que certaines catégories expressives existent, il est possible d'orienter une étude sur les aspects acoustiques présents lors des réponses à des questions d'opinions dites fermées (les choix de réponses sont prédéfinis). La première expérience a pour objectif d'identifier les variables acoustiques opérationnelles qui permettent une catégorisation objective de l'expressivité de réponses orales à des questions de sondage, surtout en ce qui a trait à la conviction. De plus, il faut s'assurer que les mesures acoustiques, qui demandent des procédures semi-automatiques, démontrent une excellente fiabilité inter-juge, ce qui sera vérifié. Le second objectif vise à observer si un sondeur ou un évaluateur peut analyser les données de façon subjective dans l'expression de la conviction. L'hypothèse est que, dans des réponses orales à des questions de sondage, des changements prosodiques présentent des aspects catégorisables reflétant un degré de conviction chez des répondants.

#### **3.2 Expérience 2**

La deuxième expérience porte sur une analyse *post facto* des réponses orales enregistrées lors d'un sondage fait par la firme *Léger* à l'occasion d'une élection municipale à Winnipeg en 2010. L'analyse vise à déterminer si une catégorisation des réponses selon les variables acoustiques associées à l'expression de la conviction, telles qu'isolées à l'Expérience 1, procure une meilleure prévision des résultats électoraux en comparaison aux réponses nominatives, car la conviction porte à l'action. L'hypothèse est qu'une nouvelle classification des réponses nominatives par rapport aux variables acoustiques exprimant la conviction diminue l'écart entre la proportion de répondants affirmant vouloir voter pour certains candidats et la proportion observée lors du vote. Les sondages se vérifient *post facto* à partir de comportements observables.

## **4. Expérience 1**

Les sections ci-dessous résument la méthodologie de la collecte des données, les analyses effectuées, les résultats et une discussion sommaire pour l'Expérience 1. L'expérience visait à identifier les variables associées à l'expression de la conviction dans des réponses orales produites dans un contexte contrôlé. Il était important d'obtenir, dans cette expérience, des réponses exprimant de façon authentique un degré de conviction. Pour ce faire, on a d'abord créé un questionnaire sur des thèmes pouvant susciter des réponses affectives reflétant les convictions des répondants. Ensuite, un groupe de participants ( $n = 25$ ) a réalisé deux tâches. La première tâche était de répondre oralement à un questionnaire enregistré portant sur des sujets d'actualités. Puis, dans une deuxième tâche, le participant devait écouter et évaluer des enregistrements de ses réponses orales selon le degré de conviction exprimée. L'objectif consistait à savoir si une décision subjective de la conviction est catégorisable et fiable. Il s'agissait d'évaluer si le jugement des individus peut servir à mesurer l'expression de la conviction. Mentionnons que les procédures et méthodes de recrutement de la présente étude ont été approuvées au préalable par le Comité d'éthique de la recherche de la Faculté des arts et des sciences (CÉRFAS).

### **4.1 Méthode**

#### **4.1.1 Participants**

Un groupe de 15 femmes et 10 hommes francophones, âgés de 18 et 44 ans, ont participé à l'expérience sur une base volontaire. Ces individus ont déclaré n'avoir aucun trouble d'audition ou de production orale diagnostiqué. Tous étaient des locuteurs natifs du français québécois. Chaque participant a signé un formulaire de consentement rattaché à une lettre d'information. Ce formulaire contenait également un questionnaire leur demandant d'indiquer leur profil sociodémographique (leur âge, leur sexe, leur langue maternelle, leur occupation et leur niveau de scolarité; voir le document en Annexe).

### 4.1.2 Stimuli

#### Élaboration d'un questionnaire pour obtenir les réactions de conviction

Dans la présente expérience, l'élaboration du questionnaire a été inspirée par le travail de Martin et de Singly (2005), qui suggèrent d'introduire les questions en demandant *pensez-vous que, trouvez-vous que, ou estimez-vous que...* Selon ces auteurs, ces formulations changent le type et le taux de réponse attendus. Dans notre expérience, on a développé un questionnaire constitué de questions et de choix de réponses au moyen d'un groupe témoin de 5 femmes et 6 hommes francophones, âgés de 18 et 44 ans. Ce groupe devait choisir dans une série de questions (58) celles qui seraient le plus susceptibles de solliciter des réactions (par exemple la conviction). Le groupe témoin était composé d'étudiants et de travailleurs avec différents niveaux de scolarités. Ces individus devaient évaluer plusieurs questions à partir d'une échelle de 1 à 10 où 1 ne signifiait « aucune réaction » et 10 une « forte réaction ». Les participants pouvaient également suggérer des questions sur des sujets non abordés. Les questions ayant reçu le plus haut taux de réactions par les juges ont été sélectionnées pour l'Expérience 1 afin d'établir le profil expressif (voir Annexe 1). Par exemple : *Êtes-vous pour ou contre la séparation du Québec?* De plus, on a ajouté quelques questions de « remplissages » (angl. *fillers*) suscitant moins de réactions. Au total, 32 questions divisées en quatre blocs (7 questions et 1 question test par bloc) ont été retenues. Les choix de réponses étaient *oui* ou *non* ; *pour* ou *contre* ; *souvent, parfois, jamais* ; *très, assez, peu, pas du tout*.

#### Enregistrement du questionnaire

Dans le cadre de cette étude, un homme, locuteur natif du français québécois, a été recruté pour enregistrer les échantillons de la voix cible (les questions et les choix de réponses). Ces échantillons ont été obtenus par l'enregistrement d'une série d'énoncés présentée sur un écran d'ordinateur dans une cabine insonorisée. Le locuteur a reçu la consigne de lire les énoncés présentés à l'écran avec un débit et un rythme naturel. En tout temps, il pouvait s'exercer et se reprendre s'il considérait que l'énoncé n'était pas naturel ou moins bien réussi (p. ex. : l'hésitation, la perte du rythme). L'enregistrement du questionnaire oral a été réalisé dans une cabine insonorisée du Laboratoire de sciences phonétiques (Université de Montréal) au moyen d'un micro canon à directivité cardioïde (Newman, modèle KMR81). Le microphone était placé à une distance constante (15 cm des lèvres et à 45

degrés environ) tout au long de l'enregistrement. L'enregistrement des énoncés a été réalisé à l'aide d'une carte de son externe de 16-bits (*Fast-track Pro*, M-Audio) combinée à un amplificateur DMP3 et réglée à un taux d'échantillonnage de 44,1 kHz, à l'aide d'un logiciel (*Goldwave*, version 5.55).

### **4.1.3 Procédures**

L'expérience comporte des procédures particulières en ce que le participant devait, dans une première tâche, répondre à un questionnaire qu'il entendait et ensuite évaluer ses réponses en écoutant un enregistrement audio. Durant ces tâches, le participant était assis dans une cabine à l'épreuve du bruit avec un micro-cravate placé à environ 6 cm des lèvres (*Audio-Technica*, modèle AT803B). Ils devaient écouter les stimuli au moyen de haut-parleurs placés dans la cabine. Les réponses des participants ont été enregistrées au moyen des mêmes caractéristiques acoustiques utilisées lors de l'enregistrement du questionnaire. À l'extérieur de la cabine, l'expérimentateur surveillait le déroulement de la séance. Un dispositif d'interphone lui permettait de communiquer avec le participant dans la cabine. Le répondant était informé qu'il y avait deux tâches.

#### **Tâche 1**

La première tâche consistait à suivre les consignes présentées à l'écran : le participant devait attendre la fin de la question et des choix de réponses avant de répondre oralement. Il devait répondre strictement selon le choix de réponse donné p. ex. *oui/non, très/assez/peu/pas du tout*, etc. afin d'éviter l'ajout d'information acoustique (lexèmes, bruits, etc.). Le participant était informé qu'il pouvait, à tout moment, refuser de répondre à une question s'il la considérait comme trop difficile ou trop personnelle. La tâche était précédée de quatre questions d'entraînement qui n'étaient pas retenues lors de l'analyse. À travers les différents blocs de questions, soit au début, au milieu et à la fin de l'expérience, le participant devait aussi lire à un débit normal les choix de réponses présentés visuellement hors contexte. Chacun des choix de réponses apparaissait au moins trois fois et dans un ordre aléatoire. L'ordre de présentation des 28 stimuli auditifs utilisés était différent pour chaque participant. La présentation des questions d'un même bloc était aléatoire et la présentation des blocs était



aléatoire pour minimiser l'effet d'habituation. Chaque question avait des choix de réponses précis.

## **Tâche 2**

Dans la deuxième tâche, le participant devait écouter ses réponses (sans le contexte) aux différentes questions de la Tâche 1. Par la suite, il devait évaluer le degré de conviction qu'il exprimait au moyen d'un questionnaire écrit (voir Annexe 1). Un autre exemplaire du même questionnaire a été remis au participant afin d'évaluer la fiabilité des jugements exprimés par une méthode de test et retest. Ce test a été appliqué sur les 350 énoncés évalués deux fois (la moitié du nombre total de réponses). Le participant ne savait pas à quelle question la réponse correspondait. Cette évaluation a été inspirée des méthodologies d'Izard 1972 (DES : *differential emotion scale*), et de Kimble et coll., 1991 sur le niveau de confiance basé sur une échelle de 1 à 7 où 1 signifie moins intense et 7 plus intense. La méthodologie d'Izard suppose que des émotions distinctes existent et sont mesurables alors que Kimble et ses collègues proposent une évaluation dans laquelle le participant indique son niveau de confiance par rapport à la réponse donnée. Par contre, avec ces méthodologies, le taux de fiabilité est faible. Cette tâche a permis d'évaluer si le participant obtenait une évaluation comparable et fiable du degré de conviction de leurs réponses (voir la section **Résultats**).

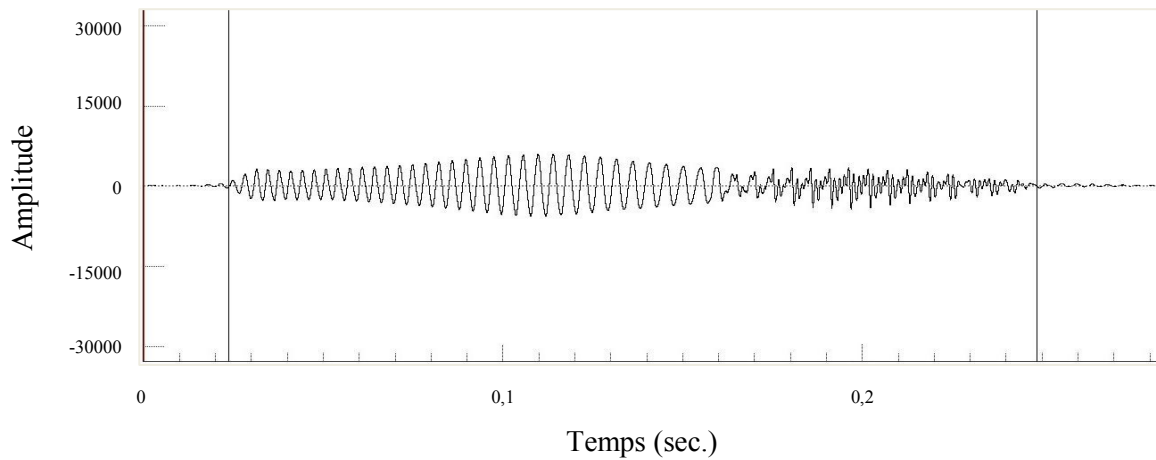
### **4.1.4 Analyse des réponses : mesures acoustiques**

Les variables dépendantes acoustiques qui ont été mesurées sont les suivantes : le débit, la durée, l'amplitude et l'intonation à partir de la fréquence fondamentale ( $F_0$ ). Il est important de noter que, pour mesurer les valeurs associées à l'expression de la conviction, il fallait définir des valeurs de référence par rapport à la parole habituelle (non affective) telle que reflétée dans la tâche de lecture des choix mis hors contexte. Cette méthode est centrale dans le présent travail étant donné les variations individuelles et contextuelles des différentes variables mesurées. Ce choix méthodologique se justifie du fait qu'il était important d'avoir une base comparative afin d'éviter tout biais dans les mesures acoustiques pouvant provenir du bruit ambiant ainsi que des variations associées à l'âge et le sexe des participants. Les exemples des variations seront abordés dans les sections portant sur la fréquence fondamentale et l'amplitude. Cette méthode de relativisation des valeurs implique une comparaison entre les

valeurs « expressives » et les valeurs « neutres ». Ce choix de comparer des réponses expressives et neutres à l'intérieur d'un même sondage permet de contrôler les facteurs externes au sondage (la fatigue, le stress du répondant), ainsi que les éléments liés à la satisfaction du sondage (intérêt du sujet, longueur du sondage, talent de l'intervieweur, etc.). Pour établir ces valeurs de référence neutres, on a fait la moyenne d'au moins trois mesures sur des réponses données hors contexte. Le choix du même lexème mis hors contexte a été utilisé, car lors d'un prétest composé de questions jugées neutres, le répondant avait la même conviction que si des questions jugées expressives avaient été utilisées. Les valeurs de références sont précisées dans chacune des variables acoustiques ci-dessous.

Les analyses portaient sur les parties des enregistrements reflétant les réponses des participants. Plusieurs fichiers ont été exclus pour différentes raisons. Les réponses peu fréquentes (« souvent » et « peu ») n'ont pas été retenues. Par ailleurs, d'autres contextes ont été éliminés du fait que la réponse relative à la valeur de référence hors contexte n'était pas neutre. Certaines mesures acoustiques étaient aussi impraticables en raison de bruits présents surtout lors de la production de consonnes. Pour éviter les données hors normes, les occlusives non voisées [p], [t], [k] et les fricatives non voisées [f], [s], [ʃ] en début de mot n'ont pas été incluses dans les mesures acoustiques. En considérant ces différentes difficultés, les mesures ont été limitées aux contextes de réponses suivants : [ɔ̃] dans *contre*; [ur] dans *pour*; [wi] dans *oui*; [nɔ̃] dans *non*; [ɛ] dans *très*; [ase] dans *assez*; [adytu] dans *pas du tout*; [arfwɑ] dans *parfois*; et [zɑmɛ] dans *jamais*. La Figure 1, ci-dessous, illustre un exemple de segmentation de réponses représentant le contexte des mesures acoustiques. La segmentation est indiquée par les deux lignes présentes avant et après le signal. Douze mesures acoustiques ont été appliquées à ce type de contexte afin d'isoler les variations pouvant s'associer aux jugements du degré de conviction d'une réponse orale. Toutes les analyses acoustiques ont été effectuées pour chaque stimulus avec le logiciel *MultiSpeech* (modèle 3700, version 3.2.1, Kay Elemetrics). Les paragraphes suivants définissent sommairement les différentes mesures.

Figure 1. Exemple d'une segmentation pour le lexème « non ».



### Fréquence fondamentale ( $F_0$ )

Les mesures de la  $F_0$  incluaient la moyenne, l'étendue (la valeur maximale moins la valeur minimale) en Hz et en semiton. On a effectué ces mesures avec une fenêtre Blackman en variant l'étendue et la progression de la fenêtre (p.ex. selon une variation de l'emplacement de la fenêtre de 25 ms et une progression de 20 ms). Une moyenne entre les différentes mesures de fenêtre a été faite. De telles mesures répétées avec différentes tailles de fenêtres et de temps de progression permettaient d'éviter les valeurs excessives ou hors-normes, généralement représentées par des points irréguliers (Figure 3) qui ne suivent pas une courbe de  $F_0$  normale (Figure 2). La  $F_0$  varie selon l'âge et le sexe de l'individu (Baken et Orlikoff, 2000). Puisqu'il y a une variation de la  $F_0$ , il faut regrouper les réponses. Pour ce faire, le même type de mesures a été appliqué à des réponses dites neutres qui ont ensuite été moyennées pour établir la  $F_0$  moyenne du participant afin de les comparer aux réponses expressives.

Figure 2. Exemple d'analyse acoustique de la fréquence fondamentale pour la réponse « non ».

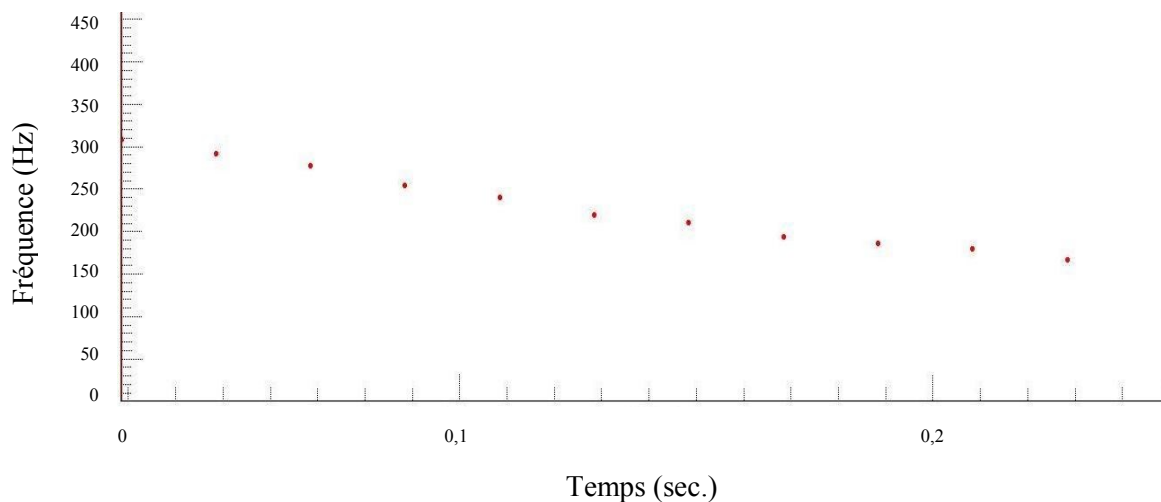
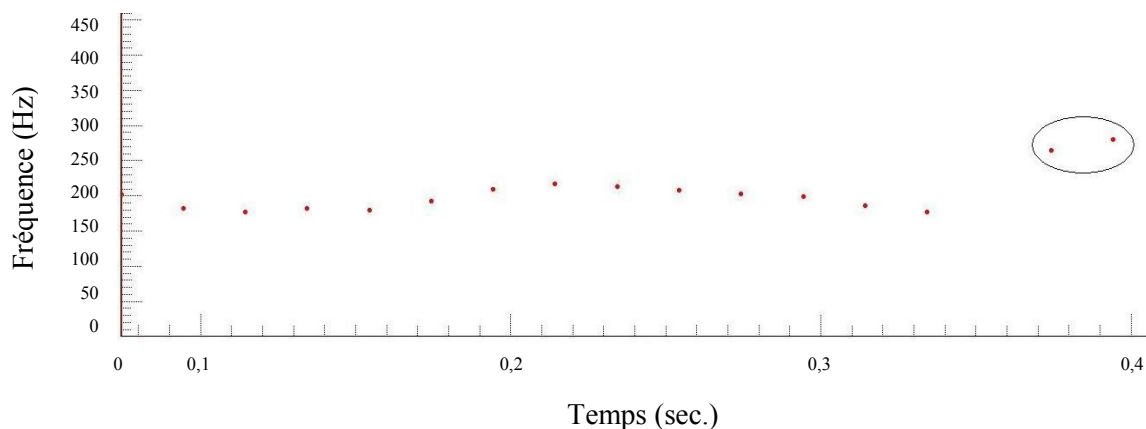


Figure 3. Exemple d'analyse acoustique de la fréquence fondamentale contenant des valeurs erronées dans la réponse « jamais ». Elles ont été encerclées.

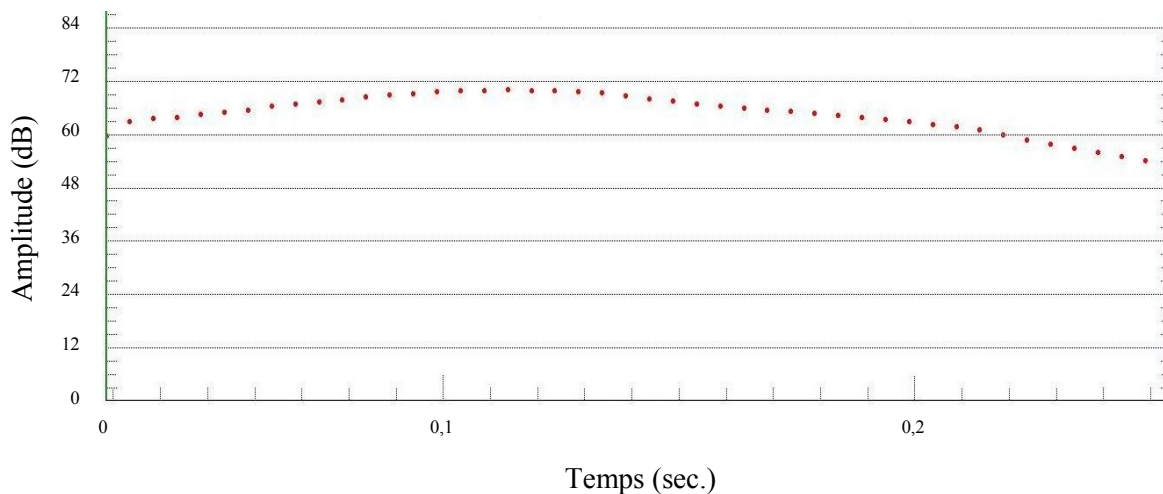


### Amplitude (en dB)

Les mesures d'amplitude ont été la moyenne, et l'étendue en décibel (dB). Dans le cas des réponses contenant plus d'une syllabe, les sommets de chaque voyelle correspondant à la valeur maximale de chaque syllabe ont aussi été considérés pour l'étendue. Une seule analyse avec une fenêtre de 5 ms et un « lissage » a été appliquée. Pour éviter l'interférence en bordure de la segmentation des signaux, la première et la dernière valeur ont été éliminées pour chaque mesure d'amplitude. Les mêmes analyses ont été appliquées sur les réponses neutres. Il est à noter que l'amplitude varie selon les locuteurs et les voyelles produites (Baken et Orlikoff,

2000). Il y a aussi des variations dans les conditions de transmission et d'enregistrement audio, par exemple, lors du placement et de la proximité du récepteur téléphonique. C'est pourquoi, par exemple, le lexème « pour » expressif devait être comparé à un autre « pour » produit en contexte neutre. La figure suivante donne un exemple d'analyse d'amplitude et de fenêtre de résultats.

Figure 4. Exemple d'analyse acoustique de l'amplitude pour la réponse « non ».



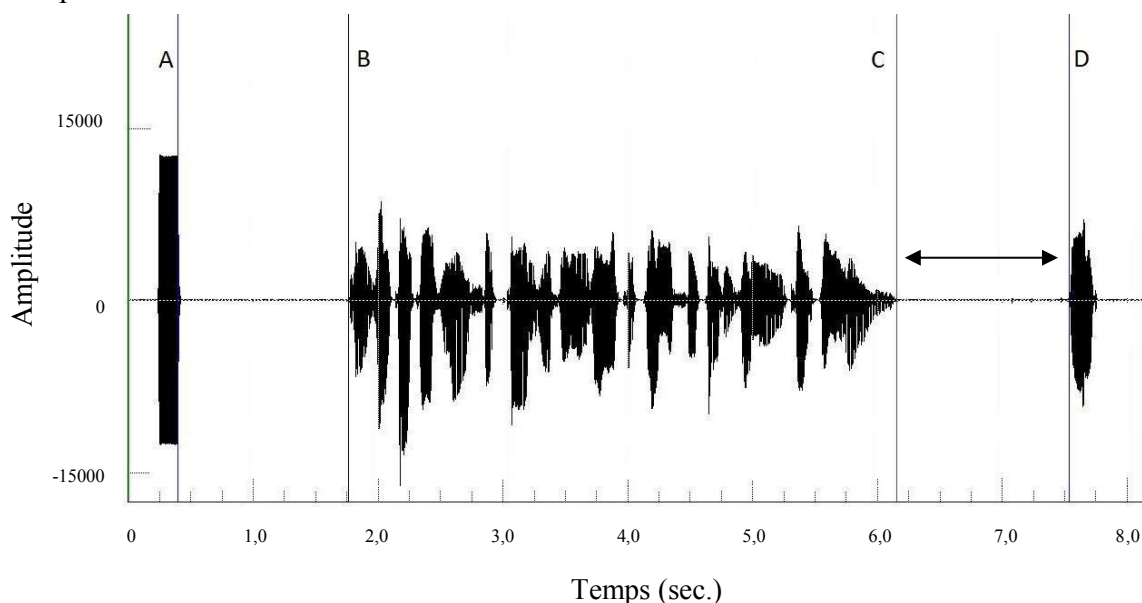
### Débit et durée

Le débit et la durée ont été mesurés à partir d'une représentation oscillographique. Le débit a été calculé en divisant le nombre de syllabes par la durée de l'énoncé (syll./durée du contexte en sec.). La durée a été mesurée (en ms) à partir de la partie périodique du lexème s'il débutait par une occlusive non voisée ([p], [t], [k]) et une fricative non voisée ([f], [s], [ʃ]). Les mêmes analyses ont été appliquées sur les réponses neutres.

### Temps de réaction

Le temps de réaction a été mesuré (en ms) à partir d'une représentation oscillographique entre la fin de la question et le début de la réponse où débute la partie périodique (voir, dans la Figure 5, l'intervalle de temps entre (C) et (D)). Cependant, il n'a pas été possible d'établir un temps de réaction neutre puisque les choix de réponses ont été dits hors contexte. La figure ci-dessous illustre un exemple de segmentation.

Figure 5. Exemple d'analyse du temps de réaction de la réponse « non ». A=bip comme indice de début de mesure; B=début de la question posée; C= fin de la question posée; D= la réponse donnée.



### Lexèmes ajoutés

Les lexèmes ajoutés ont été compilés pour les analyses exploratoires. Ces lexèmes ajoutés se situaient entre la fin de la question et le début de la réponse attendue.

#### 4.1.5 L'analyse statistique

Les analyses statistiques ont été effectuées à l'aide du logiciel *SPSS* (version 17,0). Une analyse comparative entre les réponses expressives et les réponses neutres visait à isoler les variables acoustiques qui étaient utilisées afin de rendre compte des différences dans le degré de conviction. Un simple test *t* pour échantillons appariés a été appliqué sur les variables énumérées. Les paires de mesures, dans ce cas, faisaient la comparaison entre les valeurs reflétant une réponse expressive et les valeurs d'une réponse neutre. Pour ce qui est du temps de réaction, cette variable a été analysée seulement en situation expressive (données brutes). Il est possible de voir si cette variable peut être considérée en appliquant un test *t* à échantillon unique avec un test d'hypothèse dont la valeur médiane est de zéro.

Dans le cas de l'évaluation intra-juge du degré de conviction exprimée, la fiabilité été analysée par une méthode de test et retest, telle que décrite à la section des procédures. On a

d'abord appliqué le test du coefficient de Kappa. Avec ce test, on estime qu'un  $k$  de 0,7 est acceptable. Les participants ont chacun évalué 28 énoncés selon une échelle de valeurs de 1 à 7. Deux types de catégorisation ont été testés, car l'attribution des valeurs neutres varie dans la littérature. Dans la première catégorie, les valeurs 1 à 3 reflétaient une hésitation; la valeur 4 a été interprétée en tant que neutre, alors que les valeurs 5 à 7 reflétaient une réponse ferme et convaincue. La deuxième catégorie contenait 1 et 2 pour l'hésitation; 3 à 5 pour la neutralité et 6 et 7 pour la conviction. De plus, un troisième test a été appliqué uniquement sur les données extrêmes (1-2 et 6-7) afin d'observer si ces dernières sont catégorisables de la même façon entre les essais.

## 4.2 Résultats

L'Expérience 1 avait pour objectif principal de déterminer les variables acoustiques opérationnelles ayant rapport à l'expression de la conviction à partir de mesures objectives. De plus, elle permettait d'identifier si une méthode où les répondants évaluaient des enregistrements de réponses à un questionnaire oral peut servir à mesurer l'expression de la conviction. Les résultats ci-dessous sont présentés en deux parties : la première porte sur les mesures acoustiques qui permettent d'identifier les variables pertinentes et la deuxième porte sur l'évaluation subjective.

### **Évaluation objective : les variables acoustiques qui diffèrent de la parole habituelle**

Pour évaluer la précision et la fiabilité des mesures acoustiques des réponses expressives, on a effectué des comparaisons entre les mesures obtenues par l'expérimentateur et celles produites par un juge externe. Dans ce type de vérification, on répète des mesures sur un sous-échantillon de réponses (8% dans ce cas-ci). Le juge externe connaissait les fonctionnalités du logiciel d'analyse acoustique *MultiSpeech* (modèle 3700, version 3.2.1, Kay Elemetrics) utilisé dans cette étude. Pour procéder à l'analyse du corpus, le juge externe a appliqué le même protocole de mesures décrit dans la méthode ci-dessus en utilisant les mêmes instruments. On a retiré les valeurs extrêmes à partir d'un score  $Z$ . Une comparaison des mesures a démontré une fiabilité inter-juge en ce qui a trait aux valeurs obtenues de  $F_0$  avec les différences suivantes : l'écart sur la mesure de la  $F_0$  moyenne était de  $\pm 6,34$  Hz et la différence était non significative ( $\bar{X}_1 = 172,23$ ;  $\bar{X}_2 = 172,62$ ;  $n = 28$ ;  $t = -1,08$ ; n. s.). Ensuite,

l'écart entre les juges était de  $\pm 5,23$  Hz sur la valeur minimale ( $\bar{X}_1 = 150,51$ ;  $\bar{X}_2 = 150,09$ ;  $n = 28$ ;  $t = 1,20$ ; n. s.), de  $\pm 6,96$  Hz pour la valeur maximale ( $\bar{X}_1 = 200,87$ ;  $\bar{X}_2 = 201,37$ ;  $n = 28$ ;  $t = -0,99$ ; n. s.) et de  $\pm 1,75$  sur les mesures d'étendue en semiton ( $\bar{X}_1 = 4,92$ ;  $\bar{X}_2 = 5,10$ ;  $n = 28$ ;  $t = -1,78$ ; n. s.). Quant aux valeurs correspondant à l'amplitude, les différences observées dans les mesures étaient de  $\pm 0,99$  dB pour la valeur moyenne ( $\bar{X}_1 = 61,30$ ;  $\bar{X}_2 = 61,31$ ;  $n = 28$ ;  $t = -0,08$ ; n. s.), de  $\pm 5,45$  dB pour la valeur minimale ( $\bar{X}_1 = 47,01$ ;  $\bar{X}_2 = 46,83$ ;  $n = 28$ ;  $t = 0,40$ ; n. s.) et de  $\pm 0,88$  dB pour la valeur maximale ( $\bar{X}_1 = 69,71$ ;  $\bar{X}_2 = 69,68$ ;  $n = 28$ ;  $t = 0,33$ ; n. s.). Ensuite, pour les mesures de durée, l'expérimentateur et le juge externe ont obtenu des mesures similaires avec une différence de  $\pm 15,16$  ms ( $\bar{X}_1 = 286,45$ ;  $\bar{X}_2 = 283,75$ ;  $n = 27$ ;  $t = 2,00$ ; n. s.). Enfin, les mesures pour le temps de réaction diffèrent de  $\pm 12,82$  ms ( $\bar{X}_1 = 805,53$ ;  $\bar{X}_2 = 807,35$ ;  $n = 27$ ;  $t = -1,90$ ; n. s.). En somme, les mesures acoustiques utilisées dans cette étude présentent une bonne fiabilité inter-juge.

Les résultats de l'analyse acoustique des réponses données se présentent comme suit. Un test  $t$  pour échantillons appariés a été appliqué sur les réponses neutres et expressives de chaque variable acoustique (la  $F_0$ , l'amplitude, le débit et la durée du contexte). Un test  $t$  pour échantillon unique a été appliqué sur les valeurs brutes pour le temps de réaction. La variable portant sur les lexèmes ajoutés n'a pas été considérée, car le nombre d'ajouts était faible. Les tableaux ci-dessous permettent de voir les variables acoustiques qui présentent des différences significatives par rapport à la parole neutre et qui, par conséquent, permettront d'identifier les variables qui seront utilisées pour le classement du degré de conviction exprimée.

Notons que le tableau II reflète les résultats d'analyses acoustiques des réponses pour le Bloc A, soit des réponses « oui » et « non ». À partir des données, on constate une tendance générale entre les questions pour les variables du débit, de la durée et du temps de réaction qui rendent compte des distinctions catégorielles dans l'expression de la conviction. On note que pour les réponses oui/non, le temps de réaction est un indice présent à travers différentes questions. Le tableau III présente les résultats d'analyses pour le Bloc B composé des réponses « parfois » et « jamais ». Ces réponses contiennent plus d'une syllabe. Dans ce cas,



les mesures d'étendue ont porté sur les sommets de chaque voyelle correspondant à la valeur maximale de chaque syllabe. On remarque que les variables de l'amplitude minimale et de l'étendue de l'amplitude pour chaque sommet des voyelles (sv) sont significatives dans toutes les questions du Bloc B. On voit aussi que les variables temporelles (débit, durée, temps de réaction) donnent des résultats positifs pour les réponses « parfois/jamais ». Le tableau IV présente les résultats d'analyse pour le Bloc C composé des réponses « très », « assez » et « pas du tout ». Les variables significatives sont encore une fois le débit, la durée et le temps de réaction. Les autres variables ne présentent aucune différence significative stable à travers l'ensemble des réponses. Le tableau V présente les résultats d'analyse pour le Bloc D composé des réponses « pour » et « contre ». Les variables significatives sont le débit, la durée et le temps de réaction. De plus, la réponse « pour » se distingue de façon significative par l'étendue de la  $F_0$  (en Hz et en semiton). À travers toutes les questions couvertes dans les blocs A-B-C-D, on retrouve les mêmes variables significatives : le débit, la durée et le temps de réaction. Il existe donc des variables qui, de façon constante, varient significativement par rapport à la parole neutre. Ces variables peuvent être utilisées pour distinguer la catégorisation du degré de conviction exprimée.

Tableau II. *Tests t pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc A<sup>1</sup>.*

Questions	A1	A2	A3	A4	A5	A6	A8
Réponses	Oui	Oui	Non	Non	Non	Non	Oui
	<i>n</i> = 14	<i>n</i> = 14	<i>n</i> = 19	<i>n</i> = 13	<i>n</i> = 14	<i>n</i> = 20	<i>n</i> = 16
Variables	Valeurs du test de <i>t</i>						
F <sub>0</sub>							
moy. (Hz)	1,72	2,05	-0,96	0,99	1,08	0,01	2,88*
min.	2,37*	2,40*	-1,78	0,12	-0,29	-1,03	2,15*
max.	0,19	1,07	-0,75	-0,47	0,34	-0,08	1,02
étendue (Hz)	-2,58*	-0,52	0,40	-0,87	0,49	0,87	-1,49
étendue (semiton)	-3,94**	-0,69	1,00	-0,46	0,46	1,49	-1,68
Amplitude (dB)							
moy.	-1,08	-1,39	-1,85	0,54	-0,62	-1,98	-1,19
min.	-1,35	-0,74	0,60	1,00	2,58*	0,85	0,27
max.	-0,45	-1,56	-1,23	0,97	-0,63	-1,64	-0,96
étendue	0,65	-0,63	-1,52	-0,27	-1,97	-2,07	-0,83
Débit (syll./sec.)	5,95***	3,61**	3,31**	2,95*	4,21**	4,29***	5,54***
Durée (ms)	-6,86***	-2,40*	-2,21*	-1,91	-2,72*	-3,90**	-6,38***
Temps de réaction (ms) <sup>2</sup>	4,52**	4,72***	6,69***	5,52***	4,17**	8,24***	6,36***

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Note :<sup>1</sup> : sauf le temps de réaction qui est basé sur un échantillon unique

<sup>2</sup> : sans les ajouts,  $n = 19$  pour A6

Tableau III. Tests *t* pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc B<sup>1</sup>.

Questions	B1	B2	B5	B7	B8
Réponses	Parfois <i>n</i> = 13	Parfois <i>n</i> = 14	Jamais <i>n</i> = 18	Parfois <i>n</i> = 14	Parfois <i>n</i> = 15
Variables	Valeurs du test de <i>t</i>				
F <sub>0</sub>					
moy. (Hz)	1,47	2,05	0,30	1,66	0,71
min.	-0,51	0,89	-1,78	-1,27	-0,54
max.	1,60	2,67*	0,56	2,12	1,62
étendue (Hz)	2,17	2,77*	1,64	3,20**	2,14
étendue (semiton)	2,11	2,43*	2,01	3,79**	2,18*
Amplitude (dB)					
moy.	-1,71	-2,12	-1,48	-4,24**	-3,87**
min.	-0,92	-1,46	-0,62	-1,99	-1,48
max.	-1,41	-1,50	-0,46	-0,39	-2,43*
étendue	-0,23	0,23	0,08	1,14	-0,34
min. sv <sup>a</sup>	-3,04*	-2,85*	-3,08**	-4,53**	-6,25***
max. sv <sup>a</sup>	-1,20	-1,46	-0,47	-0,25	-2,42*
étendue sv <sup>a</sup>	2,89*	2,33*	2,44*	3,96**	2,79*
Débit (syll./sec.)	5,00***	5,45***	4,85***	7,06***	7,16***
Durée (ms)	-5,05***	-5,38***	-4,85***	-8,45***	-6,85***
Temps de réaction (ms) <sup>2</sup>	8,19***	2,94*	5,02***	9,80***	5,75***

\* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001

Note : <sup>1</sup> : sauf le temps de réaction qui est basé sur un échantillon unique

<sup>2</sup> : sans les ajouts, *n* = 11 pour B2

<sup>a</sup> : valeurs des sommets des voyelles pour chaque syllabe

Tableau IV. Tests *t* pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc C<sup>1</sup>.

Questions	C4	C7	C8
Réponses	Très <i>n</i> = 20	Assez <i>n</i> = 15	Pas du tout <i>n</i> = 13
Variables	Valeurs du test de <i>t</i>		
F <sub>0</sub>			
moy. (Hz)	-0,35	1,43	-2,26*
min.	-2,42*	-1,62	-2,40*
max.	-0,44	1,84	-0,15
étendue (Hz)	2,32*	2,52*	1,30
étendue (semiton)	1,99	2,75*	1,84
Amplitude (dB)			
moy.	-1,62	-3,04**	-2,67*
min.	-0,39	-1,09	0,55
max.	-0,03	-1,78	-0,28
étendue	0,30	-0,98	-0,68
min. sv <sup>a</sup>	n/a	-3,51**	-4,10**
max. sv <sup>a</sup>	n/a	-1,72	-0,57
étendue sv <sup>a</sup>	n/a	1,53	3,34**
Débit (syll./sec.)	3,89**	5,72***	5,69***
Durée (ms)	-4,02**	-5,97***	-6,38***
Temps de réaction (ms)	9,41***	5,81***	5,45***

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Note : <sup>1</sup> : sauf le temps de réaction qui est basé sur un échantillon unique

<sup>a</sup> : valeurs des sommets des voyelles pour chaque syllabe

Tableau V. Tests *t* pour échantillons appariés des variables acoustiques en contexte expressif et neutre pour le Bloc D<sup>1</sup>.

Questions	D1	D8	D4	D5	D6	D7
Réponses	Contre	Contre	Pour	Pour	Pour	Pour
	<i>n</i> = 20	<i>n</i> = 20	<i>n</i> = 19	<i>n</i> = 19	<i>n</i> = 21	<i>n</i> = 18
Variables	Valeurs du test de <i>t</i>					
F <sub>0</sub>						
moy. (Hz)	-1,24	-0,94	-2,58*	-0,72	-0,90	-1,51
min.	-2,60*	-2,26*	-3,99**	-3,73**	-4,02**	-4,14**
max.	-0,20	0,87	0,004	2,02	2,15*	1,61
étendue (Hz)	1,70	2,37*	2,98**	5,02***	5,70***	4,10**
étendue (semiton)	1,72	2,61*	3,49**	5,40***	5,93***	4,28**
Amplitude (dB)						
moy.	-0,60	-0,91	-1,33	0,34	-1,65	-1,79
min.	-1,34	-1,60	-0,77	2,19*	0,36	-1,15
max.	-0,32	-0,80	-0,53	0,14	-1,20	-1,84
étendue	0,99	0,90	0,44	-1,73	-1,66	-0,41
Débit (syll./sec.)	3,53**	3,75**	4,28***	5,49***	4,22***	3,90**
Durée (ms)	-3,35**	-3,21**	-3,76**	-5,90***	-4,35***	-4,91***
Temps de réaction (ms) <sup>2</sup>	6,86***	10,30***	7,85***	4,64***	3,55**	6,41***

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Note : <sup>1</sup> : sauf le temps de réaction qui est basé sur un échantillon unique

<sup>2</sup> : sans les ajouts,  $n = 18$  pour D4

### **Évaluation subjective du degré de conviction exprimée**

Afin de vérifier si la catégorisation du degré de conviction peut se faire simplement par une méthode subjective, on a demandé aux participants d'évaluer le degré de conviction exprimée. On a appliqué un test de Kappa qui comparait les réponses lors d'une première évaluation et les réponses d'une deuxième évaluation. On a catégorisé les réponses de 1 à 7 selon une échelle plus catégorique où les valeurs 1-2-3 étaient combinées à la catégorie 1 (hésitation), la valeur 4 à la catégorie 2 (neutre) et les valeurs 5-6-7 à la catégorie 3 (conviction). Même avec cette nouvelle catégorisation qui favorisait un plus grand taux de correspondance, le Kappa était de 0,598, indiquant une fiabilité en deçà du seuil de 0,7. On a également réalisé un autre test de Kappa dans le cas d'une catégorisation de type 1-2, 3-4-5, 6-7 et les résultats présentaient une évaluation subjective non concordante ( $k = 0,544$ ). Un autre test a été effectué uniquement sur les valeurs extrêmes de l'échelle (soit 1-2 et 6-7). Dans ce cas, on obtient un Kappa significatif de 0,743.

### **4.3 Discussion**

En somme, les résultats répondent aux objectifs de catégorisation de l'analyse du degré de conviction exprimée dans des réponses orales à un questionnaire. Ceci est possible grâce aux variables acoustiques mesurées de façon fiable. Les variables catégorisées de façon constante à travers les questions sont le débit, la durée et le temps de réaction. Aussi, la variation de la fréquence fondamentale (étendue en Hz et en semiton) et sa valeur minimale ont aussi été significatives. Dans les cas où le lexème était composé de plus d'une syllabe, l'amplitude minimale des sommets des voyelles et leur étendue étaient significatives. Ainsi, ces variables catégoriques permettront d'indiquer qu'une réponse plus convaincue tend à refléter un temps de réaction plus rapide, une durée plus courte, une fréquence fondamentale plus variable et une amplitude plus haute et plus variable.

Pour le second objectif, qui consistait à savoir si une décision subjective est suffisamment catégorique, les résultats montrent que cette méthode d'évaluation est variable et n'est pas fiable. Donc, un sondeur utilisant uniquement cette technique n'aurait pas un degré de réponse satisfaisant, car les participants ne donnent pas toujours une note similaire pour catégoriser le niveau de conviction entendu. Il est important de mentionner que la conviction

subjective est mesurée assez régulièrement dans les sondages. Par exemple, on demande au répondant immédiatement après avoir exprimé une opinion : « Quel est votre niveau de certitude face à cette opinion ? ». Des travaux antérieurs montrent que cette conviction subjective est liée à plusieurs phénomènes qui soulignent l'importance de la conviction, dont la stabilité dans le temps des opinions (notamment: Petty & Krosnick, 1995). Ainsi, la validation subjective peut servir de complément lors de l'observation des valeurs extrême de l'échelle. Ces observations tendent à confirmer l'idée de Caelen-Haumont (2004) selon laquelle il serait préférable d'avoir une catégorisation qui peut s'établir objectivement *a posteriori* en observant les comportements à partir des données acoustiques fournies.

L'Expérience 1 a permis d'établir l'existence des variables acoustiques mesurables qui sont des indices du degré de conviction exprimée lors de réponses orales à un questionnaire utilisé lors des sondages. L'Expérience 2 exploite ces variables acoustiques afin d'établir si une quantification du degré de conviction des réponses orales permet d'améliorer l'aspect prédictif des réponses dans le cadre d'un sondage électoral.

## 5. Expérience 2

La deuxième expérience utilise une méthode expérimentale *post facto* (Silva, 2010) et a pour objectif de vérifier si les aspects acoustiques exprimant la conviction dans les réponses à un sondage téléphonique constituent des données utiles dans la prédiction des comportements. Il s'agit de vérifier dans quelle mesure un classement des réponses orales à un sondage, selon des mesures de degré de conviction, permet d'augmenter l'exactitude de prédiction du comportement d'électeurs. Dans l'expérience qui suit, on analyse certaines réponses cibles ( $n = 188$ ) obtenues lors d'un sondage téléphonique mené par la firme de recherche *Léger* à l'occasion de l'élection municipale à Winnipeg en octobre 2010. Les réponses cibles portent sur deux des quatre candidats se présentant à la mairie (Sam Katz et Judy Wasylycia-Leis). Ces deux candidats ont été retenus, car ils dominaient la compétition. Une comparaison sera présentée entre les résultats du vote, du sondage et de la nouvelle catégorisation des réponses basée sur l'analyse acoustique selon le degré de conviction. Ces comparaisons permettront de vérifier si l'analyse acoustique peut servir à raffiner les prédictions des intentions de vote par rapport à une méthodologie standard utilisée dans les sondages quantitatifs.

### 5.1 Méthode

#### 5.1.1 Participants

Dans le cadre de cette deuxième expérience, les enregistrements des réponses d'environ 400 répondants âgés d'au moins 18 ans et provenant de Winnipeg, Canada ont été remis à l'expérimentateur. Les données avaient été recueillies par la firme de recherche *Léger*, du 6 au 17 octobre 2010, dans leur centre d'appel de Montréal. Les intervieweurs avaient réalisé un total de 800 entrevues en anglais pour cette enquête. Cependant, on a reçu environ 400 fiches, car uniquement la moitié des entrevues avaient été enregistrées. Cette étude avait pour but de connaître les intentions de vote des citoyens de cette municipalité et de connaître leur opinion sur les différents enjeux et prises de décision pour leur ville. L'accès à ces données a été autorisé par la firme de sondage. L'Annexe 2 contient la lettre de consentement autorisant l'utilisation de ces appels téléphoniques. Cette deuxième expérience a également été



approuvée par le Comité d'éthique de la recherche de la Faculté des arts et des sciences (CÉRFAS).

### **5.1.2 Stimuli**

Pour des fins de confidentialité, une version partielle du questionnaire est présentée dans l'Annexe 2. La durée moyenne de l'entretien était de 7 minutes. La question faisant l'objet d'analyse dans notre étude se présentait ainsi : *If a municipal election were held today in Winnipeg, which one of these candidates would you be most likely to vote for ?* (Question Q1).

### **5.1.3 Procédures et enregistrements audio**

Les entrevues ont été réalisées par un intervieweur et un seul répondant à la fois. Au début de l'entrevue, le répondant était avisé par l'intervieweur que pour des fins de contrôle de la qualité, l'entrevue serait enregistrée (voir l'Annexe 2, à la question INT50). L'intervieweur devait poser les mêmes questions à chaque répondant et dans le même ordre. Seuls les choix de réponses variaient d'ordre de présentation. L'intervieweur devait s'assurer que le participant répondait à la totalité du sondage et aux questions sociodémographiques permettant de classer les données. En cours d'entrevue, si le participant refusait de répondre ou s'il ne connaissait pas une réponse, l'intervieweur devait reposer la question ou sinon, il acceptait la réponse *ne sait pas* ou *refus*. Les intervieweurs utilisaient la plateforme de collecte de données *Voxco Command Center* (version 5.4) pour lire les questions et enregistrer les entretiens téléphoniques. Cette plateforme permet d'enregistrer, de collecter, de traiter et d'analyser les données quantitatives (valeurs numériques indiquées par l'intervieweur pour chacune des questions). Les données numériques ont été enregistrées en format *Waveform audio file format* (.wav), échantillonné à 8 kHz et de taille 8-bits. Cependant, elles ont été gonflées à 16-bits afin d'obtenir une meilleure qualité visuelle du signal acoustique lors de l'analyse. La qualité sonore des enregistrements contenait beaucoup de bruit de fond, c'est-à-dire, du bruit provenant de l'environnement du centre d'appel et de la résidence du répondant. Il n'y a pas eu de modification apportée pour éliminer le bruit de fond, car le but était de conserver les fichiers originaux pour avoir la réalité du centre d'appel.

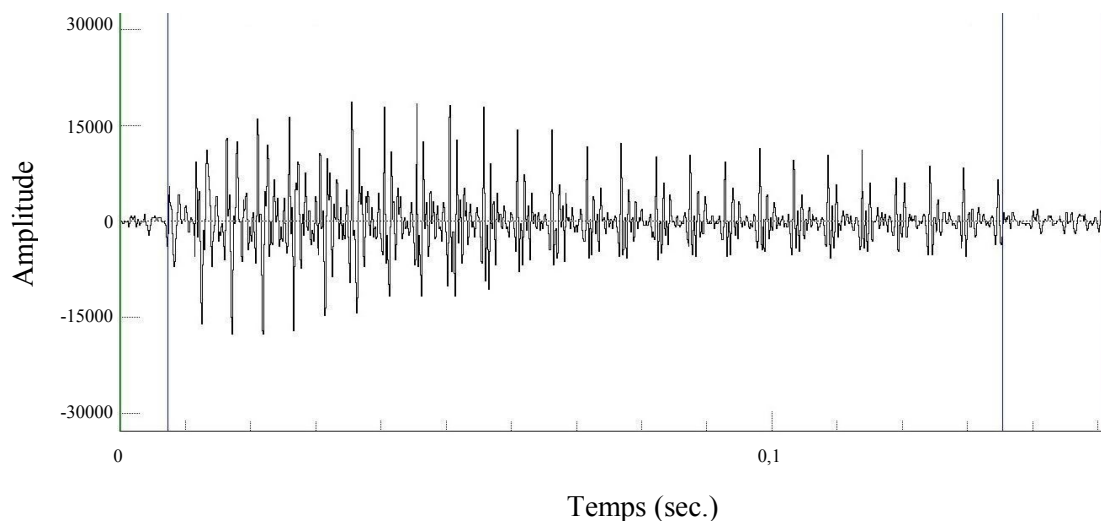
#### **5.1.4 Analyse des réponses à la question cible**

##### **Description des mesures et des variables**

À l'écoute de l'échantillon, 188 fichiers audio, correspondant à 46 % de l'échantillon reçu par la firme de recherche, ont été conservés. Ainsi, les fichiers exclus de l'analyse par l'expérimentateur l'ont été pour les raisons suivantes. Premièrement, le répondant n'était pas citoyen canadien. Dans une étude sur les tendances électorales, la citoyenneté n'est pas obligatoire, mais pour voter, elle l'est. Deuxièmement, dans les cas où les répondants n'habitaient pas la ville, ils ne pouvaient pas voter pour ces candidats. Troisièmement, si le répondant donnait une autre réponse que les choix retenus pour cette étude. Finalement, on comptait au total 283 fiches associées aux réponses « Sam » ou « Judy », de ce nombre, 34 % (95/283) de l'échantillon a été rejeté pour des raisons techniques où l'analyse acoustique était impossible. Par exemple, lorsque l'intervieweur et le répondant parlaient en même temps ou lorsqu'il y avait du bruit dans le signal.

Après vérification du corpus, les intentions de vote divisées entre les deux candidats les plus nommés et pour lesquels les résultats étaient serrés ont été retenus, c'est-à-dire « Sam Katz » et « Judy Wasylycia-Leis ». Une analyse acoustique de la voyelle [a] de « Sam » et [u] de « Judy » a été réalisée, car la voyelle est porteuse d'informations acoustiques fiables. Le [s] de « Sam » et le [dʒ] (« J ») de « Judy » n'ont pas été retenus, car ils contiennent une bande de bruit qui est difficile à mesurer et qui se confond avec le bruit de fond du centre d'appel. La figure suivante illustre un exemple de segmentation de la voyelle.

Figure 6. Exemple de segmentation de la voyelle [a] dans « Sam ».



La mesure de la voix de base de chaque participant a été prise sur la partie constante et stable de la voyelle neutre centrale [ə] présente dans les énoncés *uh*, *hum* prononcés par le participant au cours de l'entrevue. Un minimum de deux mesures de la fréquence fondamentale et de l'amplitude était requis et un calcul de la moyenne était effectué sur celles-ci. Cette mesure a été inspirée du *Speaking Fundamental frequency* de Boone, McFarlane, Von Berg, et Zraick, (2010, p. 150) qui est la  $F_0$  moyenne prise dans un discours sans interruption. Puisqu'aucune question n'était prévue pour obtenir les caractéristiques acoustiques neutres du répondant, la durée et le temps de réaction ont été mesurés à partir de la réponse brute de la question Q1 (voir l'Annexe). Ces variables ne possédaient pas de base comparative. Ainsi, les résultats reposaient uniquement sur la distribution des réponses expressives.

Les mesures acoustiques des réponses orales reflétaient les mesures faites de l'Expérience 1. En particulier, on a mesuré la fréquence fondamentale, l'amplitude, la durée et le temps de réaction de façon à évaluer si ces variables acoustiques peuvent être utilisées pour catégoriser les réponses. Les réponses neutres et expressives ont été comparées. À la différence de l'Expérience 1, des valeurs de classement différenciées entre les réponses neutres et expressives ont été utilisées étant donné les variations individuelles et contextuelles des différentes variables mesurées. Cette méthode nécessite un calcul où les valeurs

« expressives » sont soustraites aux valeurs « neutres ». Ensuite, ces résultats ont été utilisés en comparaison avec les résultats réels et ceux produits par la firme de recherche. Dans cette expérience, il est à noter que plusieurs réponses ont été fournies par des personnes âgées. En effet, leur voix affecte la fréquence fondamentale, car leur intonation est plus basse que la moyenne. La durée a été mesurée au lieu du débit, car il n'y avait qu'une seule syllabe dans tous les cas. La mesure pour le temps de réaction a été faite dès que l'intervieweur a terminé la lecture jusqu'à la réponse du nom du candidat (« Sam », « Judy »), plus précisément, jusqu'à la partie périodique de la voyelle. La mesure du temps de réaction s'est effectuée uniquement dans les cas où l'on ne retrouvait pas l'ajout de lexèmes. Sinon, il se crée un effet confondant affectant le temps de réaction, car on constate des ajouts de toute sorte (abordé plus en détail dans la section **Résultats**).

### **L'analyse statistique**

L'analyse statistique a porté sur des comparaisons simples dans SPSS. Un test *t* pour échantillons appariés a été appliqué sur les variables de l'intonation et de l'amplitude pour savoir quelles variables permettent une catégorisation. Les paires étaient composées, pour une variable, de sa valeur de base et de sa valeur expressive. Un test *t* pour échantillon unique a été appliqué sur les variables de durée et du temps de réaction. On a sélectionné des valeurs de coupes (angl. *cut-offs*) qui catégorisent le degré de conviction pour cette réponse à partir de l'observation d'histogramme. Les histogrammes illustrent, par variable et par candidat, le nombre de répondants classés selon la différence entre leur réponse à la Q1 et leur réponse neutre. Chaque variable affiche une séparation entre deux catégories pour le regroupement des réponses orales.

## **5.2 Résultats**

L'Expérience 1 a servi à identifier les variables acoustiques opérationnelles associées à l'expression de la conviction. Partant de ces premières observations, l'Expérience 2 avait deux objectifs. Le premier était d'évaluer si les variables opérationnelles lors des réponses orales enregistrées à partir d'un sondage électoral pouvaient être utilisées. Cet objectif était d'obtenir également une catégorisation d'expressivité dans les réponses de façon à les classer selon un

degré de conviction (les résultats pour ce premier objectif sont présentés aux sections 5.2.1 et 5.2.2 ci-dessous). Le deuxième objectif était de valider ce classement des réponses en examinant les gains possibles dans la prédiction des résultats des élections (ces résultats sont présentés à la section 5.2.3 ci-dessous). Essentiellement, on voulait établir si un reclassement des réponses nominatives (où les répondants désignaient un candidat préféré), selon les variables acoustiques de conviction, permettait une plus grande exactitude prédictive des résultats. Toutefois, une étape préalable à ces analyses était de vérifier la fiabilité des mesures acoustiques elles-mêmes.

Pour vérifier le degré de fiabilité des mesures acoustiques, on a effectué des tests de fiabilité identiques à ceux de l'Expérience 1 (voir la section 4.2) où l'on compare les résultats de mesures obtenues par l'expérimentateur à ceux faits par un juge externe sur un ensemble de réponses. Dans ce type de vérification, on répète des mesures sur un sous-échantillon de réponses (8% dans ce cas-ci). Le juge externe connaissait les fonctionnalités du logiciel d'analyse acoustique *MultiSpeech* (modèle 3700, version 3.2.1, Kay Elemetrics) utilisé dans cette étude. Pour procéder à l'analyse du corpus, le juge externe a appliqué le même protocole de mesures décrit dans la méthode ci-dessus en utilisant les mêmes instruments. Une comparaison des mesures a démontré une fiabilité inter-juge en ce qui a trait aux valeurs obtenues de  $F_0$  avec les différences suivantes : l'écart sur la mesure de la  $F_0$  moyenne était de  $\pm 5,21$  Hz et la différence était non significative ( $\bar{x}_1 = 150,64$ ;  $\bar{x}_2 = 150,17$ ;  $n = 12$ ;  $t = 0,98$ ; n. s.). L'écart entre les juges était de  $\pm 13,86$  Hz sur la valeur minimale ( $\bar{x}_1 = 139,70$ ;  $\bar{x}_2 = 138,67$ ;  $n = 12$ ;  $t = 0,82$ ; n. s.), de  $\pm 2,41$  Hz pour la valeur maximale ( $\bar{x}_1 = 163,99$ ;  $\bar{x}_2 = 163,86$ ;  $n = 12$ ;  $t = 0,39$ ; n. s.) et de  $\pm 1,75$  sur les mesures d'étendue en semiton ( $\bar{x}_1 = 2,44$ ;  $\bar{x}_2 = 2,78$ ;  $n = 12$ ;  $t = -1,94$ ; n. s.). Quant aux valeurs relatives à l'amplitude, les différences observées dans les mesures étaient de  $\pm 0,65$  dB pour la valeur moyenne ( $\bar{x}_1 = 69,82$ ;  $\bar{x}_2 = 69,94$ ;  $n = 12$ ;  $t = -1,41$ ; n. s.), de  $\pm 3,46$  dB pour la valeur minimale ( $\bar{x}_1 = 64,92$ ;  $\bar{x}_2 = 65,10$ ;  $n = 12$ ;  $t = -0,59$ ; n. s.) et de  $\pm 0,37$  dB pour la valeur maximale ( $\bar{x}_1 = 71,93$ ;  $\bar{x}_2 = 71,98$ ;  $n = 12$ ;  $t = -1,14$ ; n. s.). Enfin, pour les mesures de la durée, l'expérimentateur et le juge externe ont obtenu des mesures similaires avec une différence de  $\pm 10,57$  ms ( $\bar{x}_1 = 131,26$ ;  $\bar{x}_2 = 129,94$ ;  $n = 12$ ;  $t = 1,02$ ; n. s.). Les mesures pour le temps de

réaction n'ont pas été comparées, car il n'y avait pas suffisamment de mesures sans les lexèmes ajoutés. En somme, de façon générale, les mesures acoustiques utilisées dans cette étude présentent une fiabilité inter-juge.

### **5.2.1 Les variables servant au classement**

Les résultats de l'analyse acoustique de la voyelle [a] de « Sam » et [u] de « Judy » se présentent comme suit. D'une part, on devait observer si les catégories étaient comparables à l'Expérience 1. Un test  $t$  pour échantillons appariés a été appliqué sur les valeurs expressives et neutres pour chaque variable acoustique de l'intonation et de l'amplitude. Un test  $t$  pour échantillon unique a été appliqué sur les valeurs brutes pour chaque variable acoustique de la durée et du temps de réaction. Les résultats complets du test, affichant toutes les variables, se retrouvent dans le tableau ci-dessous.

Tableau VI. Statistiques et tests *t* pour échantillons appariés des variables acoustiques en contexte expressif et neutre<sup>1</sup>.

Variables	Sam		<i>n</i> = 102		Judy		<i>n</i> = 86	
	Expressif	Neutre	<i>t</i>		Expressif	Neutre	<i>t</i>	
F <sub>0</sub> (Hz)								
moy.	$\bar{x}$	164,7	166,5	- ,9	167,5	167,7	- ,9	
	<i>é.t.</i>	(45,0)	(41,0)		(44,7)	(41,8)		
min.	$\bar{x}$	153,0	163,6	-5,5***	154,2	164,4	-3,7***	
	<i>é.t.</i>	(41,4)	(41,0)		(39,9)	(41,1)		
max.	$\bar{x}$	181,7	169,3	5,3***	181,1	171,9	2,9**	
	<i>é.t.</i>	(51,0)	(41,3)		(50,4)	(43,0)		
étendue (Hz)	$\bar{x}$	28,7	5,7	17,0***	26,9	7,5	13,6***	
	<i>é.t.</i>	(13,8)	(1,6)		(14,1)	(3,5)		
étendue (semiton)	$\bar{x}$	2,9	,6	19,9***	2,6	,8	15,2***	
	<i>é.t.</i>	(1,0)	(,3)		(1,1)	(,3)		
Amplitude (dB)								
moy.	$\bar{x}$	68,6	70,2	-5,1***	70,4	70,3	,4	
	<i>é.t.</i>	(2,8)	(1,2)		(3,3)	(1,3)		
min.	$\bar{x}$	65,1	69,6	-13,8***	65,6	69,5	-8,8***	
	<i>é.t.</i>	(3,1)	(1,2)		(3,8)	(1,3)		
max.	$\bar{x}$	71,0	70,7	,8	72,5	71,0	3,7***	
	<i>é.t.</i>	(3,1)	(1,2)		(3,6)	(1,4)		
étendue	$\bar{x}$	5,9	1,0	16,0***	6,8	1,5	14,7***	
	<i>é.t.</i>	(3,0)	(,4)		(3,3)	(,7)		
Durée (ms)	$\bar{x}$	124,0	n/a	37,4***	118,2	n/a	33,8***	
	<i>é.t.</i>	(33,5)			(32,5)			
Temps de réaction (ms) <sup>2</sup>	$\bar{x}$	745,8	n/a	11,5***	758,1	n/a	10,3***	
	<i>é.t.</i>	(423,3)			(366,9)			

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Note : <sup>1</sup> : sauf la durée et le temps de réaction qui sont basés sur un échantillon unique

<sup>2</sup> : sans ajout,  $n = 43$  pour Sam et  $n = 25$  pour Judy

À partir des données présentes dans ce tableau, il en ressort que la plupart des variables rendent compte des distinctions catégorielles. Plus précisément, à la réponse « Sam », les variables qui obtiennent la plus grande valeur au test *t* appariés sont, dans l'ordre, l'étendue de la F<sub>0</sub> en semiton [ $t(101) = 19,9$ ;  $p < .001$ ], l'étendue en Hz [ $t(101) = 17,0$ ;  $p < .001$ ] et l'étendue de l'amplitude [ $t(101) = 16,0$ ;  $p < .001$ ]. Elles fournissent des catégories significativement différentes par rapport aux réponses neutres et elles permettent de rendre compte des catégories sur la conviction. La durée [ $t(101) = 37,4$ ;  $p < .001$ ] et le temps de réaction

[ $t(42) = 11,5; p <.001$ ] fournissent également des différences significatives. Il a été impossible d'avoir une référence neutre pour la durée et le temps de réaction, car ils n'en ont pas fourni. Étant donné qu'il n'y avait pas de question prévue pour mesurer la neutralité d'une réponse, il était difficile de trouver une durée et un temps de réaction correspondant à la neutralité sans influencer les résultats. Alors on a appliqué un test pour échantillon unique sur les réponses expressives. Conséquemment, les variables ont été analysées en comparant les réponses expressives brutes à partir d'un test  $t$  à échantillon unique. Alors, il n'est pas possible de le comparer avec le neutre pour voir des différences. Ensuite, à la réponse « Judy », les variables qui obtiennent la plus grande valeur au test  $t$  appariés sont, l'étendue de la  $F_0$  en semiton [ $t(85) = 15,2; p <.001$ ], l'étendue en Hz [ $t(85) = 13,6; p <.001$ ] et l'étendue de l'amplitude [ $t(85) = 14,7; p <.001$ ]. La durée [ $t(85) = 33,8; p <.001$ ] et le temps de réaction [ $t(24) = 10,3; p <.001$ ] fournissent également des différences significatives. Ainsi, la catégorisation est comparable entre les deux types de réponses et ces variables peuvent être utilisées comme valeurs de classement. On a également observé les lexèmes ajoutés. Plusieurs lexèmes indiquaient clairement le niveau de conviction (p.ex. *I'm not sure* ou *I'll vote for...*). Cependant, il y avait une grande variabilité des lexèmes utilisés pour les considérer. Il en est également pour ceux étant difficiles à catégoriser. Dans certains cas, il était difficile de juger si c'était propre à une hésitation ou à une conviction. Certains lexèmes liés à l'hésitation, telle que *hum*, *uh*, pouvaient être utilisé malgré une réponse exprimant la conviction. Donc, l'analyse des ajouts est subjective et il est difficile d'établir sa catégorie.

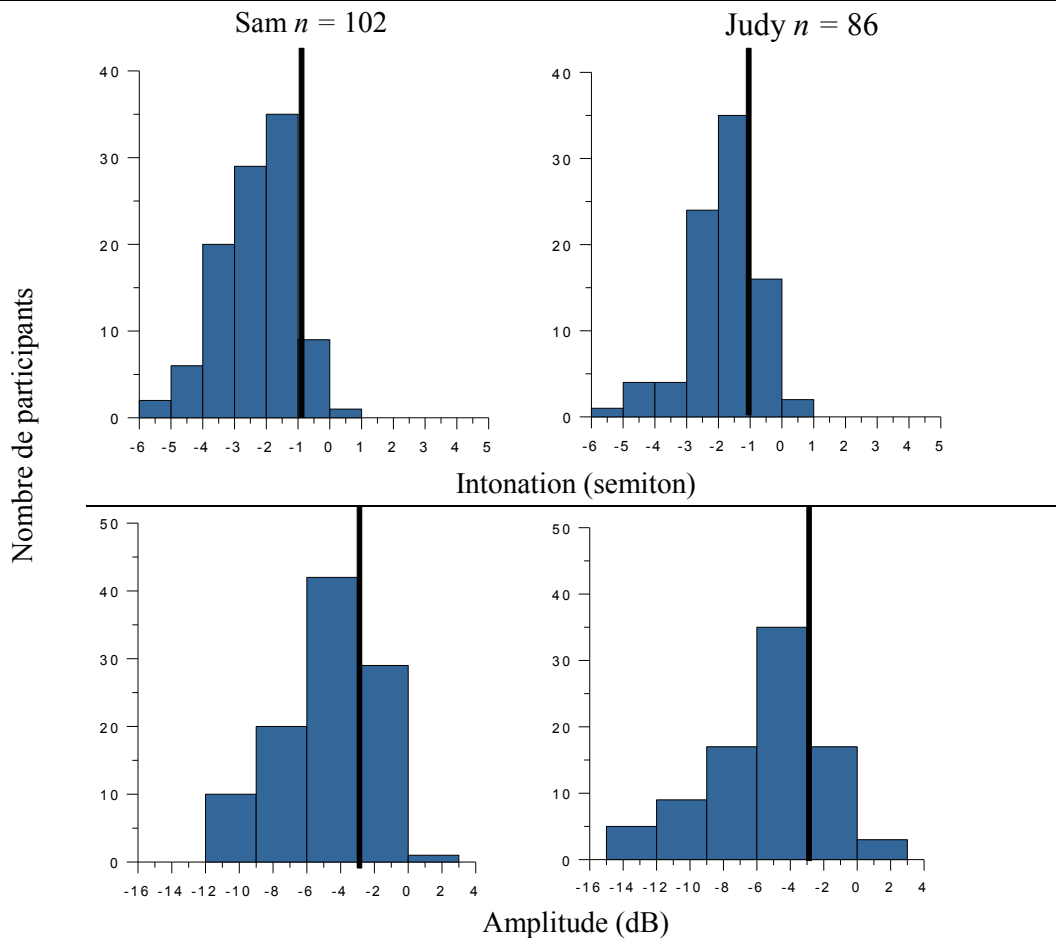
L'Expérience 2 se compare à l'expérience 1 pour les variables retenues. En observant uniquement les réponses, à une syllabe, on constate que l'étendue de la  $F_0$  en semiton et en Hz, la durée et le temps de réaction ont été des variables significatives, tout comme dans l'Expérience 1. Cependant, la variable de l'étendue de l'amplitude n'a pas été significative pour l'Expérience 1 contrairement à l'Expérience 2.



### 5.2.2 Les valeurs de classement

La distribution des valeurs acoustiques observées permet d'inférer des valeurs *de coupe* pouvant servir, dans un premier temps, à classer les réponses exprimant un degré de conviction dans le choix oral des candidats. Dans la plupart des cas, de simples histogrammes présentent des distributions où des sauts brusques ou des pointes suggèrent des valeurs de catégorisation. Les variables retenues pour le classement sont l'étendue de l'intonation de la  $F_0$  en semiton, l'étendue de l'amplitude, la durée et le temps de réaction puisqu'elles ont été significatives pour les deux réponses et qu'elles sont relatives à la conviction selon la littérature. Dans la Figure 7, on remarque que l'intonation a une frontière qui se situe à environ 1 semiton de différentiel avec sa valeur de référence. Étant donné que les valeurs sont basées sur un différentiel entre la valeur neutre et la valeur expressive, plus elle est négative (-4, -3, -2), plus elle indique une plus grande variabilité de l'intonation et une réaction du répondant. Donc sur cette base, toutes les réponses avec une  $F_0$  en semiton et ayant un écart de plus de 1 seraient classées comme ayant une plus grande variabilité expressive pour la conviction. Dans cette même figure, on remarque que l'amplitude a une frontière à 3 dB pour les deux réponses et est aussi basée sur un différentiel entre les valeurs neutres et expressives. Donc sur cette base, toutes les réponses ayant une amplitude dont l'écart est plus grand de 3 dB seraient classées comme étant plus convaincues.

Figure 7. *Histogrammes illustrant la distribution des valeurs différenciées observées pour l'intonation et l'amplitude et les valeurs de coupe<sup>1</sup> estimées servant à catégoriser un répondant selon sa conviction.*



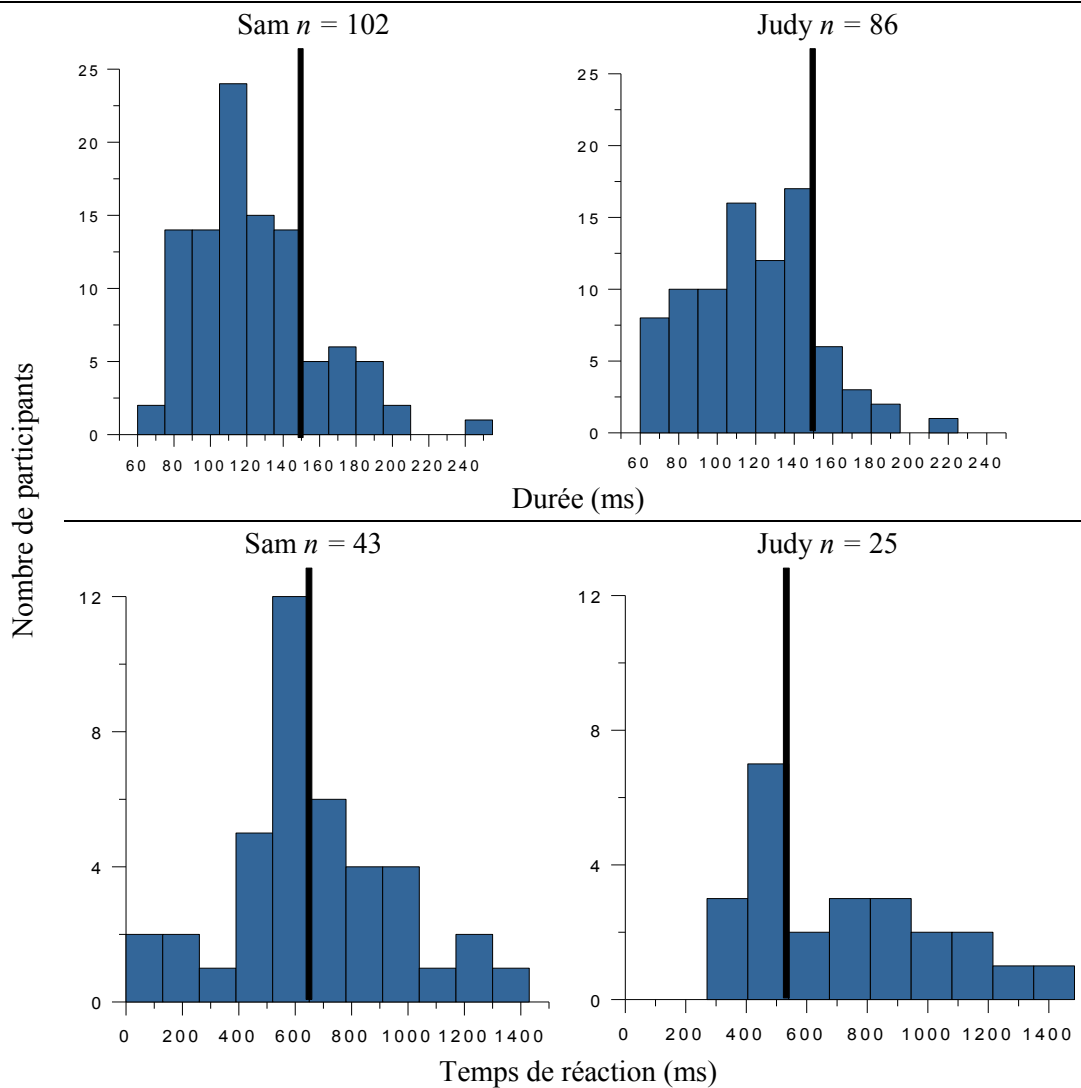
Note : Les valeurs négatives indiquent la différence entre la valeur de référence (neutre) et la valeur expressive, 0 tend vers le neutre (pour intonation et amplitude)

<sup>1</sup> : Valeurs de coupe : lignes noires servant à distinguer les catégories

Par la suite, la Figure 8 regroupe les variables correspondant à la durée et au temps de réaction. Dans le cas de la durée, l'histogramme affiche une séparation entre deux catégories pour le regroupement des réponses orales. Ces histogrammes distinguent ceux qui ont répondu le plus rapidement par rapport à ceux qui ont répondu le plus lentement. Pour le temps de réaction, une séparation a été effectuée parmi les fiches ne contenant pas d'ajout. Les lexèmes ajoutés n'ont pas été analysés compte tenu de leur variabilité, car on y retrouvait différents ajouts qui produisaient un effet confondant sur le temps de réaction. Par exemple, on retrouvait des discussions, des questionnements, des informations précisant sa conviction ou

son hésitation. Par contre, le tout était d'une durée variable et ne pouvait être utilisé pour des fins d'analyse. Les histogrammes pour la variable de la durée et du temps de réaction ont été basés uniquement sur les réponses expressives, sans comparaison avec une valeur de référence. À partir des histogrammes, on constate que la durée de la voyelle se situe à 150 ms et que le temps de réaction se situe en moyenne à 600 ms. Donc, sur cette base, toutes les réponses ayant une durée plus courte que 150 ms et un temps de réaction plus rapide qu'environ 600 ms seraient classées comme étant plus convaincues. Puisque le nombre de répondants n'était pas assez élevé pour le temps de réaction (19 participants ont répondu plus rapidement « Sam » et seulement 10 pour « Judy »), cette variable ne pourra être utilisée dans la validation des variables de classement.

Figure 8. Histogrammes illustrant la distribution des valeurs brutes observées pour la durée et le temps de réaction<sup>1</sup> et les valeurs de coupe<sup>2</sup> estimées servant à catégoriser un répondant selon sa conviction.



Note : <sup>1</sup> : Le temps de réaction avec les réponses sans ajout

<sup>2</sup> : Valeurs de coupe : les lignes noires servent à distinguer les catégories

### 5.2.3 La validation des variables de classement

Les valeurs de classement permettent d'identifier les frontières propres aux réactions expressives associées à la conviction. Ainsi, ces valeurs sont transposées en pourcentage et appliquées comme résultats pouvant être comparés aux résultats réels et à ceux du sondage. Pour obtenir ces pourcentages de comparaison, les histogrammes présentant les frontières illustrent également le nombre de répondants qui ont produit la réponse selon le profil

acoustique associé à la conviction. Le tableau comparatif ci-dessous regroupe les résultats réels de l'élection, ceux des fichiers retenus provenant du sondage de la firme *Léger* et ceux obtenus à partir des réponses provenant de l'analyse acoustique.

Tableau VII. *Tableau comparatif des résultats bruts des sondages et des analyses acoustiques sur les élections municipales à Winnipeg, 2010.*

Candidats	Résultats électoraux	Résultats du sondage*	Résultats acoustiques* <sup>1</sup>		
			Étendue de l'intonation	Étendue de l'amplitude	Durée
Sam	% 56,1	54,3	57,5	52,2	52,9
Katz	(n) (116308)	(102)	(92)	(72)	(83)
Judy	% 43,9	45,7	42,5	47,8	47,1
Wasylycia-Leis	(n) (90913)	(86)	(68)	(66)	(74)
<b>Total</b>	<b>n = 207221</b>	<b>n = 188</b>	<b>n = 160</b>	<b>n = 138</b>	<b>n = 157</b>

\*Données partielles et retenues provenant de la firme *Léger*, qui réalisa des entrevues téléphoniques entre le 6 et le 17 oct. 2010.

<sup>1</sup> : Résultats provenant des variables acoustiques catégorisées et illustrées par les histogrammes ci-haut.

Le nombre de répondants correspond au nombre de personnes faisant partie de la catégorie expressive recherchée. Le nombre de personnes affiché varie en fonction de l'analyse acoustique effectuée. Par exemple, 83 personnes ont donné une réponse plus courte lorsqu'ils ont donné la réponse « Sam », comparativement à 74 personnes pour « Judy ».

Les principaux résultats illustrent que les variables acoustiques mesurables comme la variabilité de l'intonation, la variabilité de l'amplitude et la durée poursuivent un lien logique à la conviction exprimée. La variable de l'étendue de l'intonation en semiton se rapproche des résultats, car son pourcentage pour la réponse « Sam » est de 57,5 comparativement à 56,1 pour les résultats réels et pour la réponse « Judy » de 42,5 par rapport à 43,9. Les variables présentent des « seuils » applicables servant à classer le degré de conviction.

### 5.3 Discussion

Les résultats suggèrent que la conviction peut être différenciée à partir de mesures sur un signal acoustique. Les analyses présentées à la section précédente démontrent la relation entre les variables acoustiques et la conviction basée sur des observations entre la voix neutre et la voix expressive du répondant. En effet, cette étude compare les résultats réels de l'élection et les réponses littérales du sondage avec les données acoustiques. L'étude exploratoire avait pour but de maximiser les résultats à partir des informations acoustiques. Les catégories étaient établies *post facto* afin d'obtenir les meilleurs résultats. En ce sens, nous ne pouvions pas prévoir les écarts. À partir des corrélats acoustiques, on atteint presque les mêmes résultats, et ce, en utilisant moins d'échantillons. Les principales variables acoustiques sont la variation de l'intonation (par la fréquence fondamentale), la variation de l'amplitude et la durée. Le temps de réaction était significatif, par contre il aurait fallu plus de participants ayant répondu uniquement les choix de réponse demandés. La présence de plusieurs variables donne l'avantage que si un participant est rejeté pour une cause technique durant l'analyse, il est possible que d'autres variables soient utilisées pour ne pas perdre le répondant. De plus, les principaux résultats de cette étude indiquent que les variables acoustiques en lien avec la conviction servent à prédire le comportement électoral. Elles peuvent être utilisées en tant qu'outil complémentaire à l'analyse des données.

On sait que l'analyse acoustique fonctionne. Toutefois, pour valider cette étude de façon empirique, on se doit de connaître le vote réel des répondants en réalisant une étude postélectorale. Cependant, la firme de recherche n'a pas recontacté ces mêmes répondants pour connaître leur vote. Cette précision permettrait d'appuyer la prédiction, car elle est nécessaire pour réaliser une analyse logistique régressive, cette dernière permet de prédire l'appartenance d'une réponse à un groupe prédéfinie de façon objective.

Pour être en mesure d'appliquer ces résultats à d'autres enquêtes téléphoniques, il faut apporter quelques recommandations afin d'utiliser des mesures acoustiques. Lors de la création du questionnaire, il faudrait avoir des questions qui permettent au répondant de produire un discours « neutre » afin d'obtenir des mesures de références pour l'amplitude et l'intonation. En début de sondage, il faut donner des instructions claires au répondant. Par

exemple, on doit l'informer qu'il doit attendre que la question soit terminée pour répondre et ainsi de ne pas interrompre l'intervieweur. Il est important de bien former les intervieweurs sur l'obtention de réponses expressives. Pour ce faire, il faut s'assurer qu'ils lisent les questions au complet, et n'interrompent pas le répondant. Ensuite, il faut vérifier quelques aspects techniques concernant l'enregistrement sonore. Il faudrait demander au répondant de garder une distance constante entre sa bouche et le microphone du téléphone. Et faire en sorte d'éviter les bruits environnants qui nuisent à la qualité de l'enregistrement. Une autre recommandation serait d'enregistrer les entrevues sur deux chaînes audio, ainsi si l'intervieweur parle en même temps que le répondant, les données pourront être utilisées.

## 6. Discussion générale et conclusion

Les deux expériences ci-dessus permettent de constater que certaines mesures acoustiques servent à catégoriser objectivement la conviction exprimée et qu'il est possible d'utiliser des valeurs ou des catégories de valeurs pour prédire les intentions de vote. Rappelons que l'Expérience 1 avait pour objectif principal d'identifier les variables acoustiques opérationnelles qui permettent une catégorisation de l'expressivité de différentes réponses orales à des questions de sondage. Le second objectif de l'Expérience 1 était de voir si, indépendamment de mesures objectives, des individus pouvaient classer les réponses de façon subjective, selon un degré de conviction. L'idée était de voir si le jugement d'un sondeur, à lui seul, pouvait servir à classer des réponses entendues de façon fiable. Ces résultats ont orienté l'Expérience 2 qui avait pour objectif de déterminer l'applicabilité d'une catégorisation des réponses selon des mesures acoustiques objectives.

Pour résumer les résultats de l'Expérience 1, rappelons que cette expérience reposait sur l'hypothèse que des changements mesurables dans la voix des répondants présentent des aspects catégoriques en lien avec le degré de conviction exprimée. Afin d'examiner cette possibilité, on a appliqué différentes mesures acoustiques (de la fréquence fondamentale, de l'amplitude, du débit et du temps de réaction) sur les réponses à un questionnaire fermé portant sur des sujets d'actualité. Le choix des questions permettait de solliciter des réponses avec différent degré de conviction. Par exemple, on posait des questions sur la séparation du Québec ou encore sur la grève en tant que moyen de pression. Ces questions ont permis d'obtenir des réponses exprimant l'hésitation, la neutralité et la conviction. Nos mesures acoustiques effectuées sur ces réponses ont démontré que plusieurs variables acoustiques permettent de la catégoriser. En particulier, comme illustrées dans le Tableau V, les variables significatives de débit, de durée et de temps de réaction permettent une catégorisation. On a aussi vu que l'analyse de la réponse « pour » se distingue de façon significative par l'étendue de la  $F_0$  (en Hz et en semiton). Par ailleurs, les résultats de cette première expérience ont aussi servi à clarifier le fait que, bien qu'il y ait des variables exprimant la conviction, l'évaluation subjective ne peut servir à identifier de façon fiable des catégories d'expressivité. En somme,



les résultats suggéraient qu'il existe bel et bien des variables objectives permettant de classifier les réponses orales à un sondage téléphonique selon l'expression de la conviction, mais que ces variables ne pouvaient pas être classées de façon fiable par une écoute subjective. Ce mémoire mène à certaines interrogations sur l'utilisation conjointe de la conviction révélée par le répondant et de la conviction acoustique.

En référence à ces premiers résultats, l'Expérience 2 visait à déterminer si une classification objective des réponses orales à un sondage, selon un degré de conviction, permettait d'améliorer les prédictions des résultats électoraux. Dans cette optique, on a utilisé les données réelles d'un sondage fait par la firme *Léger* à l'occasion d'une élection municipale à Winnipeg en 2010. Les données indiquent que la reclassification des réponses, selon les mesures acoustiques reflétant un degré de conviction, permet de diminuer l'écart ou d'augmenter la prévision des résultats électoraux en comparaison aux réponses littérales. Ceci a permis de distinguer l'écart entre la proportion de répondants affirmant vouloir voter pour certains candidats et la proportion de votes obtenus par un candidat. En somme, cette expérience suggère que la nouvelle catégorisation selon des valeurs acoustiques associées à l'expression de la conviction pourrait améliorer la méthodologie existante du sondage afin d'obtenir des prédictions plus exactes des intentions de vote. Notons que certaines études précédentes utilisant des données de sondage ont fait usage de variables de durée, de temps de réaction et de lexèmes ajoutés (Schaeffer et coll, 1993; Schober et coll, 2004). Toutefois, dans ces études, les données sur l'intonation et l'amplitude n'avaient pas été recueillies lors d'un sondage d'opinion. Or elles sont nécessaires pour identifier la conviction (voir le Tableau VI). Aussi, la grande différence entre cette étude et les autres études est que celle-ci a permis de prédire les résultats d'une élection.

Pris ensemble, les résultats des deux expériences démontrent qu'il est possible de catégoriser objectivement les informations expressives des réponses orales à partir de variables acoustiques. Ainsi, les réponses orales contiennent des informations acoustiques sur le degré de conviction des répondants et ces informations peuvent servir à prédire des comportements. Dans cette perspective, bien qu'un sondage d'opinion soit axé sur des réponses de types littérales (« oui/non »; nom des politiciens, etc.), des informations portant

sur l'affectivité des répondants en lien avec la conviction ou l'hésitation du locuteur ont une importance dans la précision des prédictions issues des sondages. Les résultats obtenus à partir d'un sondage prédictif, comparable à une réalité observable, renforcent l'utilisation des mesures acoustiques comme outil d'analyse de la conviction pour tous les types de réponse et de réactions. Ces informations, considérées « qualitatives », ne sont pas habituellement considérées dans les sondages de types quantitatifs, car les firmes emploient déjà d'autres méthodes pour obtenir ces valeurs (p. ex. dans les groupes de discussion). Notons aussi que les informations concernant des mesures acoustiques paraissent difficiles à obtenir. En fait, dans notre étude, l'analyse de mesures a été faite manuellement et a pris plusieurs heures. Toutefois, il est important de noter que plusieurs variables se mesurent au moyen de techniques automatisées. Cela est le cas des mesures de la fréquence fondamentale, de l'amplitude, du débit et du temps de réaction, où des mesures automatiques pourraient s'opérer et assurer une fiabilité dans les valeurs extraites. Toutefois, d'autres mesures, telles que les lexèmes ajoutés, doivent être identifiées manuellement. Ainsi, les résultats du présent travail indiquent les avantages et les possibilités de procédures automatiques de détection de la conviction dans la voix.

## 7. Bibliographie

- Bachorowski, J. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, 8(2), 53-57.
- Baken, R. J. (1987). *Clinical measurement of speech and voice*. Boston: College-Hill Press. 518p.
- Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice* (2nd Ed.). San Diego: Singular Thomson Learning. 604p.
- Balconi, M., Brambilla, E., & Falbo, L. (2009). Appetitive vs. defensive responses to emotional cues. Autonomic measures and brain oscillation modulation. *Brain Research*, 72-84.
- Beller, G. (2007). Influence de l'expressivité sur le degré d'articulation. *RJCP, Rencontres Jeunes Chercheurs de la Parole*, 24-27.
- Boone, D. R., McFarlane, S. C., Von Berg & S. L., Zraick, R. I. (2010). *The voice and voice therapy* (8e ed.). Boston, MA: Allyn & Bacon. 360p.
- Brennan, S., & Williams, M. (1995). The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*, 34(3), 383-398.
- Caelen-Haumont, G. (2004). Les états émotionnels et la Prosodie: paradigmes, modèles, paramètres. *Phonologie et phonétique: Forme et substance*, 397-424.
- Cohen, A., Hong, L., & Guevara, A. (2009). Understanding emotional expression using prosodic analysis of natural speech: Refining the methodology.
- Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition et emotion*, 27(4), 658-684.
- Cowie, R., & Cornelius, R. (2003). Describing the emotional states that are expressed in speech. *Speech communication*, 40(1-2), 5-32.
- Ekman, P. (1982). *Emotion in the human face*. New York: Cambridge University Press. 439p.
- Greco, V., et Roger, D. (2003). Uncertainty, stress, and health. *Personality and Individual Differences*, 34(6), 1057-1068.

- Guilbert, P., Haziza, D., Ruiz-Gazen, A., & Tillé, Y. (2008). *Méthodes de sondage : enquêtes électorales, enquêtes dans le domaine de la santé, enquêtes dans les pays en développement : cours et cas pratiques*. Paris: Dunod. 380p.
- Gustafson-Capková, S. (2001). Emotions in Speech: Tagset and Acoustic Correlates. *Speech Technology*, 1-13.
- Izard, C. (1972). *Patterns of emotions: A new analysis of anxiety and depression*: Academic Press New York. 301p.
- Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. dans J. A. Harrigan, R. Rosenthal et K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65-136). Oxford, New York: Oxford University Press.
- Kimble, C., & Seidel, S. (1991). Vocal signs of confidence. *Journal of Nonverbal Behavior*, 15(2), 99-105.
- Larousse. *Larousse.fr : encyclopédie et dictionnaires gratuits en ligne. Édition Internet*, [En ligne], 2014. [<http://www.larousse.fr/>] (4 novembre 2013)
- LeDoux, J. (1996). *The emotional brain : the mysterious underpinnings of emotional life*, New York : Simon et Schuster. 384p.
- Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughhead, J., Valdez, J. N., Javitt, D. N., Turetsky, & B. I., Gur, R. C. (2010). "It's Not What You Say, But How You Say It": a reciprocal temporo-frontal network for affective prosody. *Frontiers in human neuroscience*, 4.
- Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27(4), 363-384.
- Martin, O., & de Singly, F. (2005). *L'enquête et ses méthodes : L'analyse de données quantitatives*. Paris: Nathan Université. 126p.
- Monat, A., Averill, J. R., & Lazarus, R. S. (1972). Anticipatory stress and coping reactions under various conditions of uncertainty. *Journal of personality and social psychology*, 24(2), 237.
- Mozziconacci, S. (2002). *Prosody and emotions*. Paper presented at the Speech Prosody, Aix-en-Provence, France.
- Murray, I., & Arnott, J. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal Acoustical Society of America*, 93, 1097-1097.

- Nygaard, L., Herold, D., & Namy, L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33(1), 127-146.
- Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols et J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-365). Cambridge: Cambridge University Press.
- Pantic, M., Sebe, N., Cohn, J., & Huang, T. (2005). *Affective multimodal human-computer interaction*.
- Petty, R. E., & Krosnick, J. A. (1995). *Attitude Strength: Antecedents and Consequences*. Mahwah, NJ: Lawrence Erlbaum Associates. 510p.
- Pickett, J. M. (1999). *The acoustics of speech communication: Fundamentals, speech perception theory, and technology*. Boston: Allyn and Bacon. 404p.
- Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International journal of psychophysiology*, 61(1), 5-18.
- Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, 29(5), 483-495.
- Russell, J., Bachorowski, J., & Fernandez-Dols, J. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54(1), 329-349.
- Schaeffer, N., Maynard, D., & Cradock, R. (1993). Negotiating certainty: Uncertainty proposals and their disposal in standardized survey interviews. *Center for Demography and Ecology Working Paper*.
- Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99(2), 143-165.
- Scherer, K., London, H., & Wolf, J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, 7(1), 31-44.
- Schober, M., & Bloom, J. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse processes*, 38(3), 287-308.
- Schröder, M. (2001). *Emotional speech synthesis: A review*. Paper presented at the Eurospeech, Aalborg, Denmark.
- Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. C. (2001). *Acoustic correlates of emotion dimensions in view of speech synthesis*. Paper presented at the INTERSPEECH.

- Silva, C. (2010). Ex Post Facto Study. In Neil J. Salkind (Ed.), *Encyclopedia of Research Design*. (pp. 466-467). Thousand Oaks, CA: SAGE Publications, Inc.
- Vaissière, J. (2004). Les universaux de substance prosodiques. *Sophie Wauquier (éd.), Les universaux sonores. Rennes: Presses Universitaires de Rennes.*
- Vaissière, J. (2006). Perception of intonation. Dans D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception*. Malden, MA: Blackwell.
- Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, 52, 1238.

# Annexe 1 : documents pour l'Expérience 1

## Questions utilisées en laboratoire afin de recueillir les réactions

Pratique 1 (A\_7) Le Canada doit continuer à payer pour les procès de terroristes.

Pratique 2 (B\_4) C'est plus rentable de donner des services uniquement en anglais au Québec.

Pratique 3 (C\_5) L'installation d'un compteur d'eau à votre domicile pour que l'eau vous soit facturée selon la quantité que vous utilisez.

Pratique 4 (D\_2) L'hypersexualisation des femmes dans les publicités.

### **BLOC A : choix *Oui/Non***

A\_1 L'éducation devrait être gratuite.

A\_2 Il faut interdire aux femmes musulmanes le droit de voter en portant une burqa qui cache leur visage.

A\_3 L'armée canadienne serait dans ses droits d'envahir le Québec en cas de séparation.

A\_4 Les attentats du 11 septembre 2001 étaient planifiés par les Américains eux-mêmes.

A\_5 Tout le monde devrait parler anglais au Québec.

A\_6 La loi 101 est la cause de la tuerie à Dawson.

A\_8 Il faut obliger les immigrants à apprendre le français en premier.

### **BLOC B : choix *Souvent/Parfois/Jamais***

B\_1 Je fais des gestes concrets pour sauver la planète.

B\_2 J'ai déjà dit des propos racistes.

B\_3 Les Québécois travaillent moins que les Ontariens.

B\_5 Comme l'a déclaré le Dr Mailloux, les Noirs ont un quotient intellectuel plus faible que la moyenne.

B\_6 Il faut une meilleure sélection très restrictive de l'immigration au Canada.

B\_7 Je donne de l'argent aux mendiants.

B\_8 Faire la grève est un bon moyen de pression.

### **BLOC C : choix *Très/Assez/Peu/Pas du tout***

C\_1 J'aime les émissions de télé-réalité.

C\_2 Il faut limiter la vitesse des voitures à 120 km/h.

C\_3 Les hommes auront toujours un meilleur salaire que les femmes.

C\_4 Toute personne gravement blessée a droit à l'aide médicale.

- C\_6 La présence de l'armée canadienne en Afghanistan est nécessaire.
- C\_7 Obama va réussir à modifier la perception des Américains dans le monde.
- C\_8 Il faut avoir des services uniquement en anglais au Québec.

**BLOC D : choix *Pour/Contre***

- D\_1 La position de l'église qui interdit le droit d'avortement chez les femmes violées.
- D\_3 La séparation du Québec.
- D\_4 Le suicide assisté pour les personnes souffrantes.
- D\_5 Une réévaluation de la conduite automobile dès l'âge de 65 ans.
- D\_6 L'installation de vélo libre-service payant sur l'île de Montréal.
- D\_7 Le retour d'une équipe de hockey à Québec.
- D\_8 Le retrait des cours d'arts dans les écoles secondaires.



## Tâche 2, auto-évaluation subjective

En tenant compte de la manière dont vous avez donné vos réponses, veuillez indiquer quelle était l'intensité de vos réactions (degré de conviction, certitude, etc.). Encerchez le chiffre qui s'applique.

1.1	- Intense	1	2	3	4	5	6	7	+ Intense
1.2	- Intense	1	2	3	4	5	6	7	+ Intense
1.3	- Intense	1	2	3	4	5	6	7	+ Intense
1.4	- Intense	1	2	3	4	5	6	7	+ Intense
1.5	- Intense	1	2	3	4	5	6	7	+ Intense
1.6	- Intense	1	2	3	4	5	6	7	+ Intense
1.7	- Intense	1	2	3	4	5	6	7	+ Intense
<b>2.1</b>	<b>- Intense</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>+ Intense</b>
2.2	- Intense	1	2	3	4	5	6	7	+ Intense
2.3	- Intense	1	2	3	4	5	6	7	+ Intense
2.4	- Intense	1	2	3	4	5	6	7	+ Intense
2.5	- Intense	1	2	3	4	5	6	7	+ Intense
2.6	- Intense	1	2	3	4	5	6	7	+ Intense
2.7	- Intense	1	2	3	4	5	6	7	+ Intense
<b>3.1</b>	<b>- Intense</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>+ Intense</b>
3.2	- Intense	1	2	3	4	5	6	7	+ Intense
3.3	- Intense	1	2	3	4	5	6	7	+ Intense
3.4	- Intense	1	2	3	4	5	6	7	+ Intense
3.5	- Intense	1	2	3	4	5	6	7	+ Intense
3.6	- Intense	1	2	3	4	5	6	7	+ Intense
3.7	- Intense	1	2	3	4	5	6	7	+ Intense
<b>4.1</b>	<b>- Intense</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>+ Intense</b>
4.2	- Intense	1	2	3	4	5	6	7	+ Intense
4.3	- Intense	1	2	3	4	5	6	7	+ Intense
4.4	- Intense	1	2	3	4	5	6	7	+ Intense
4.5	- Intense	1	2	3	4	5	6	7	+ Intense
4.6	- Intense	1	2	3	4	5	6	7	+ Intense
4.7	- Intense	1	2	3	4	5	6	7	+ Intense

## Formulaire de consentement

Les signatures et les informations personnelles ont été retirées.



**Laboratoire de sciences phonétiques**  
Département de linguistique et de traduction

### **Catégorisation de l'expressivité orale : étude pour un raffinement des données de sondage téléphonique.**

Je déclare avoir pris connaissance des informations ci-dessus, avoir obtenu les réponses à mes questions sur ma participation à l'étude et comprendre le but, la nature, les avantages, les risques et les inconvénients de cette étude.

Après réflexion, je consens librement à prendre part à cette recherche. Je sais que je peux me retirer en tout temps sans préjudice et sans devoir justifier ma décision.

Signature : \_\_\_\_\_ Date : \_\_\_\_\_

Nom : \_\_\_\_\_ Prénom : \_\_\_\_\_

Je déclare avoir expliqué le but, la nature, les avantages, les risques et les inconvénients de l'étude et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature de l'étudiant chercheur \_\_\_\_\_ Date : \_\_\_\_\_

Nom : \_\_\_\_\_ Prénom : \_\_\_\_\_

Pour toute question relative à la recherche, vous pouvez communiquer avec mon directeur de recherche au numéro de téléphone indiqué ci-dessous ou à l'adresse courriel suivante : \_\_\_\_\_

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal. (L'ombudsman accepte les appels à frais virés).

**Professeur responsable :**  
Victor J. Boucher

**Étudiante-chercheuse :**  
Anne-Marie Lespérance

**Un exemplaire signé du formulaire ci-dessus doit être remis au participant**

---

À quel groupe d'âge appartenez-vous ?  18-24 ans  25-34 ans  35-44 ans

Sexe :  Homme  Femme

Langue maternelle : \_\_\_\_\_

Autre(s) langue(s) parlée(s) à la maison : \_\_\_\_\_

Occupation principale : \_\_\_\_\_

Niveau de scolarité :  Primaire  Secondaire  Collégial  Universitaire

---

## Annexe 2 : documents pour l'Expérience 2

### Lettre de consentement

Les signatures et les informations personnelles ont été retirées.



Laboratoire de sciences phonétiques  
Département de linguistique et de traduction

### Étude pour un raffinement des données de sondage téléphonique : partie 2

Ce formulaire stipule que la firme de recherche *Léger Marketing* autorise à Anne-Marie Lespérance, étudiante chercheuse à la maîtrise, et à Victor Boucher, professeur responsable à l'Université de Montréal, l'accès à des enregistrements audio d'entrevues d'enquête réalisées par *Léger Marketing* à Montréal. Dans cette étude, les répondants étaient informés que l'entrevue était enregistrée et le consentement était signifié par leur consentement à participer. La firme de recherche *Léger Marketing* autorise l'accès à Anne-Marie Lespérance et le professeur Victor Boucher aux seules informations contenues dans les enregistrements, qui sont anonymes et confidentielles. Les enregistrements demeurent la propriété de *Léger Marketing* et le nom de l'entreprise devra être cité dans toute publication issue de l'étude susmentionnée. De plus, une copie de toute publication devra être soumise au Vice-président Recherche de *Léger Marketing* et les copies d'enregistrements devront être détruites trois ans suivant la fin de l'étude.

Je Anne-Marie Lespérance, déclare avoir expliqué le but, la nature, les avantages, les risques et les inconvénients de l'étude au Vice-président Recherche de *Léger Marketing*, M. Christian Bourque, pour obtenir cette autorisation d'accès.

\_\_\_\_\_  
**Professeur responsable**  
Victor J. Boucher

\_\_\_\_\_  
**Étudiante chercheuse**  
Anne-Marie Lespérance

\_\_\_\_\_  
**Vice-président *Léger Marketing***  
Christian Bourque

Une copie conforme de ce formulaire a été remise au Vice-président *Léger Marketing*

---

## **Questionnaire partiel de la firme *Léger***

### **INT50**

Before starting the interview, I would like to inform you that for quality control reasons, this interview may be recorded. May we begin the interview?

### **Q1**

If a municipal election were held today in Winnipeg, which one of these declared mayoral candidates would you be most likely to vote for? (Read list)

[Rav Gill (Rav Gill)] [01] - [Brad Gross (Brad Grawss)] [02] - [Sam Katz (Sam Kaytz)] [03] - [Judy Wasylycia Leis (Judy Wahsa-lisha Lease)] [04] - [DO NOT READ - None of the above/ I don't plan to vote] [97] - [DO NOT READ - Don't know / Undecided] [98] - [DO NOT READ - Refused] [99]

### **AGE**

For classification purposes, I now need to ask you a few additional questions.

Can you tell me in which age group you belong? Is it between...?

[...18-24 years of age] [1] - [...25-34 years of age] [2] - [...35-44 years of age] [3] - [...45-54 years of age] [4] - [...55-64 years of age] [5] - [or 65 years of age or older] [6] - [Refusal] [9]

### **OCCUP**

What is your current main occupation?

[OFFICE WORKER] [01] - [PERSONNEL SPECIALIZED IN SALES] [02] - [PERSONNEL SPECIALIZED IN SERVICES] [03] - [MANUAL WORKERS] [04] - [SKILLED, SEMI-SKILLED WORKERS] [05] - [SCIENCE AND TECHNOLOGIES WORKERS][06]-[PROFESSIONALS][07][MANAGERS/ADMINISTRATORS/OWNERS] [08] - [HOMEMAKER] [09] - [STUDENT] [10] - [RETIRED] [11] - [UNEMPLOYED] [12] - [Other] [96]O - [Refusal] [99]

### **ENFAN**

Are there any children who are UNDER 18 YEARS OF AGE living at home? (IF SO): Are any of them 12 YEARS OLD AND OLDER or LESS THAN 12 years old?

[Yes: 12 years of age and older AND younger than 12 years old] [2]

[Yes, 12 years of age and OLDER ONLY] [1] - [Yes: YOUNGER than 12 years old ONLY] [3] [No children under 18 years old at all in the household] [4]- [Refusal] [9]

**SCOL**

What is the last year of schooling that you have completed? Is it...

[... Elementary (7 years or less)] [1] - [... High School, general or professional (8 to 12 years)] [2] - [... College pre-university, technical training, certificate (CEP), accreditation (ASP) or proficiency diploma (DEP) (13-15 years)] [3] - [... University certificates and diplomas] [4] - [... University Bachelor (including classical studies)] [5] - [... University Master's] [6] - [... University Doctorate (PhD)] [7] - [Refusal] [9]

**REVEN**

Among the following categories, which one best reflects your total INCOME, before taxes, of all the members of your household in 2009? Is it...

[...\$19,999 or less] [1] - [...between \$20,000 and \$39,999] [2] - [...between \$40,000 and \$59,999] [3] - [...between \$60,000 and \$79,999] [4] - [...between \$80,000 and \$99,999] [5] - [or \$100,000 or more] [6] - [Refusal] [9]

**LANGU**

What is the language you first learned at home in your childhood and that you still understand?

[French] [1] - [English] [2] - [Other] [3] - [French and English] [7] - [French and other] [4] - [English and other] [5] - [Other and other] [6] - [Dnk/Refusal] [9]

**SEXE**

[RECORD GENDER OF RESPONDENT]

[Male] [1] - [Female] [2]

**INT99**

LEGER MARKETING THANKS YOU FOR YOUR PARTICIPATION

(By completing the interview ("CO"), the interviewer attests: "I declare that this interview was conducted with the person mentioned, in conformity with the generated sample and that all the questions requiring a response were duly provided by the respondent.")