Université de Montréal

L'effet de la familiarité sur l'identification des locuteurs :
Pour un perfectionnement de la parade vocale

par
Julien Plante-Hébert

Département de Linguistique et Traduction
Faculté des Arts et Sciences

Mémoire présenté à la Faculté des Arts et Sciences
en vue de l'obtention du grade de Maîtrise en linguistique

Août 2014

Résumé et mots clés

La présente étude porte sur les effets de la familiarité dans l'identification d'individus en situation de parade vocale. La parade vocale est une technique inspirée d'une procédure paralégale d'identification visuelle d'individus. Elle consiste en la présentation de plusieurs voix avec des aspects acoustiques similaires définis selon des critères reconnus dans la littérature. L'objectif principal de la présente étude était de déterminer si la familiarité d'une voix dans une parade vocale peut donner un haut taux d'identification correcte (> 99 %) de locuteurs. Cette étude est la première à quantifier le critère de familiarité entre l'identificateur et une personne associée à « une voix-cible » selon quatre paramètres liés aux contacts (communications) entre les individus, soit *la récence* du contact (à quand remonte la dernière rencontre avec l'individu), *la durée* et *la fréquence* moyenne du contact et *la période* pendant laquelle avaient lieu les contacts.

Trois différentes parades vocales ont été élaborées, chacune contenant 10 voix d'hommes incluant une voix-cible pouvant être très familière; ce degré de familiarité a été établi selon un questionnaire. Les participants (identificateurs, n = 44) ont été sélectionnés selon leur niveau de familiarité avec la voix-cible. Toutes les voix étaient celles de locuteurs natifs du franco-québécois et toutes avaient des fréquences fondamentales moyennes similaires à la voix-cible (à un semi-ton près). Aussi, chaque parade vocale contenait des énoncés variant en longueur selon un nombre donné de syllabes (1, 4, 10, 18 syll.). Les résultats démontrent qu'en contrôlant le degré de familiarité et avec un énoncé de 4 syllabes ou plus, on obtient un taux d'identification avec une probabilité exacte d'erreur de $p < 1 \times 10^{-12}$. Ces taux d'identification dépassent ceux obtenus actuellement avec des systèmes automatisés.

Mots clés : parades vocales, identification de locuteurs, identification par la voix, légal, longueur des énoncés, nasalité.

Abstract and keywords

The present study deals with the effects of familiarity on speaker identification in the context of voice line-ups. The voice line-up is a paralegal technique, inspired by a visual identification procedure. The voice line-up consists in presenting a number of voices sharing similar acoustic parameters as specified in established procedures. The main objective was to determine if the familiarity of a voice could lead to a high rate of correct identification (> 99 %). Our study is the first to quantify the familiarity criterion linking an identifier and a « target voice ». The quantification was based on four parameters bearing on the degree of contact between individuals: *recency, frenquency*, *duration*, and the *period* during which the contact occurred.

Three different voice line-ups were elaborated, each containing 10 voices, including one target voice which was well known by the identifier according to a questionnaire that served to quantify familiarity. Participants (identifiers, n = 44) were selected on the basis of their familiarity with the target voice. The speakers used in the voice line-ups were native speakers of Quebec French and all presented voices had similar fundamental frequencies (to within one semitone). In each line-up we used utterances of 4 different lengths (1, 4, 10, and 18 syll.). The results show that by controlling the familiarity criterion, a correct identification rate of a 100 % is obtained with an exact error probability of $p < 1 \times 10^{-12}$.These rates are superior to current automatic systems of voice identification.

Keywords : voice line-ups, speaker identification, voice identification, forensic, utterance length, nasal

Table des matières

Liste des tableaux

Liste des figures

Listes des sigles et abréviations

dB, dBa	Décibel, décibel pondéré
F0, F0mp	Fréquence fondamentale, fréquence fondamentale moyenne de la parole
G	Groupe de participants
Hz, kHz	Herz, kilo Herz
j	Participant
ms.	Milliseconde
n	Nombre de participants
N	Nombre d'essais
p	Probabilité d'erreur associée à un test statistique
P	Probabilité d'erreur globale; probabilité exacte
s.	Seconde
syll.	Syllabe
X	Essai
Z	Somme des essais
θ	Paramètre de surdispersion (voir Yu et Zelterman (2002))
Σ	Est élément de

Remerciements

Merci aux membres du Laboratoire de sciences phonétiques de l'Université de Montréal, au Pr. Karim Oualkacha du département de mathématiques de l'UQAM ainsi qu'à mes collègues du département de linguistique pour leur aide et leur collaboration.

Merci également à ma famille, tout particulièrement mes parents, et à mes amis pour le précieux soutien, les encouragements et l'aide offerte.

1 Introduction

La capacité de tout être humain à identifier un individu par le biais de sa voix peut sembler une habileté commune. Toutefois, il s'agit d'une capacité spécifique et largement inexpliquée. En effet, l'être humain est en mesure de saisir l'information présente dans le signal de la parole avec tant de précision qu'il arrive à identifier sans l'ombre d'un doute la voix d'un proche parmi une multitude de voix entendues. Cette capacité d'identification peut sembler banale, mais elle surpasse actuellement les systèmes automatisés (Jain, Flynn, et Ross, 2007; Neustein et Patil, 2012; Unar, Seng, et Abbasi, 2014). C'est-à-dire, bien que les technologies de pointe arrivent à retirer une grande quantité d'information de la parole, l'humain est toujours le seul à pouvoir identifier des voix connues avec une précision approchant la certitude. En soi, cette capacité a des implications théoriques fondamentales, mais elle présente aussi plusieurs possibilités d'applications, dont certaines dans le domaine judiciaire.

En particulier, l'identification d'un locuteur dans un contexte légal, souvent en matière criminelle, demeure un défi pour plusieurs professionnels du droit. Tel que soutenu par Interpol (2001), l'une des techniques les plus adaptées à ce jour est l'utilisation de la *parade vocale*, technique fondée sur l'identification du locuteur par un individu.

> *«Today, procedures in speaker identification by witnesses for evidential purposes typically involve the use of line-ups, following existing practice in the related domain of visual identification of persons by witnesses […]. It is worth stressing that the use of single person identification procedures, while producing positive identification scores in controlled experiments that are comparable to those obtained for (multi-person) line-ups, is generally rejected except to confirm an earlier identification. The reason is that line-ups, unlike identification procedures involving a single speaker only (i.e., the suspect), make it possible to detect the vast majority of false identifications. Detection is possible because in a properly designed line-up all members should have an equal probability of false identification, which reduces the risk of a false identification going undetected to 1/N in an N-person lineup, where 1/N is the likelihood of a false identification involving an innocent suspect. By contrast, there is of course no way in which false-positive identifications can be distinguished from correct identifications if only a single speaker is presented to the witness: both correct and incorrect identifications amount to selection of the suspect». (Interpol 2001p. D2-57)*

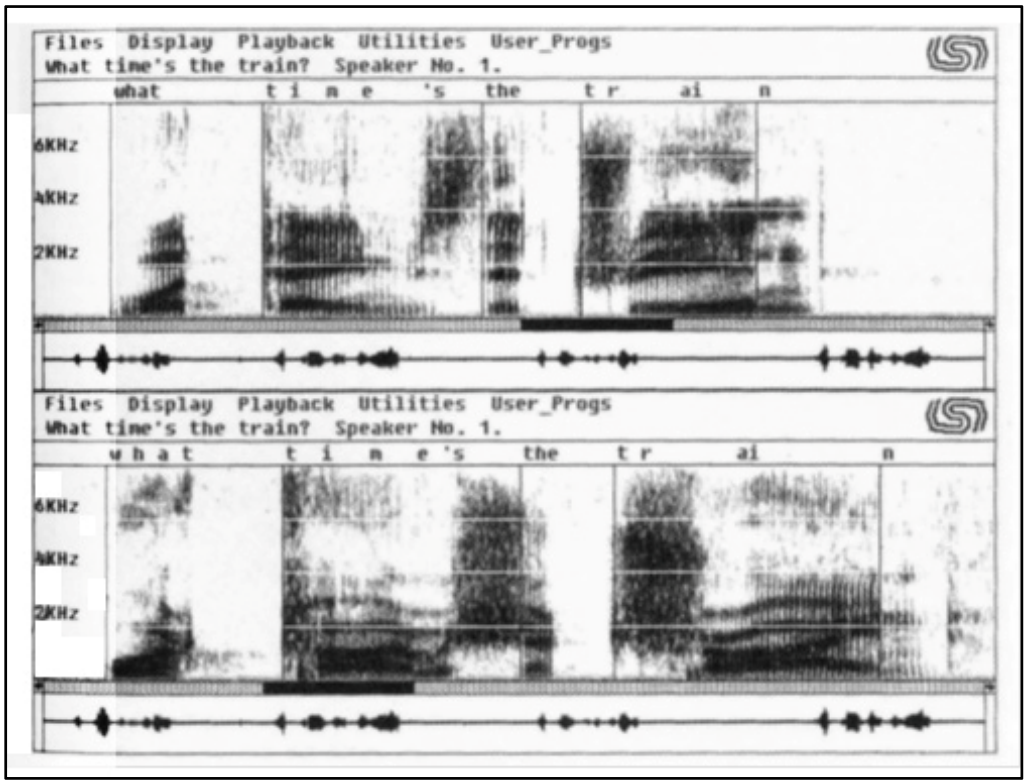
Cette citation d'Interpol met en évidence l'avantage qu'ont les parades vocales de permettre la détection d'une grande majorité de fausses identifications. On mentionne que la possibilité de fausses identifications dans une parade vocale élaborée adéquatement est de 1/*N* (*N* = nombre de voix présentes dans la parade). En pratique, la parade vocale se veut une analogie de la parade visuelle, technique bien connue des enquêteurs (Jessen, 2008; Philippon, Cherryman, Bull, et Vrij, 2007b). La parade visuelle consiste à faire défiler un groupe d'individus partageant des traits physiques semblables devant un témoin. Parmi ces individus figure un suspect au centre d'une enquête. Le témoin doit alors identifier l'individu qu'il reconnaît comme ayant commis l'infraction. Par rapport à cette technique, la parade vocale représente une procédure analogue, mais qui est centrée sur la voix d'individus plutôt que sur leurs attributs physiques. La sélection des voix dans une parade ne se fait donc pas selon des caractéristiques physiques des individus, mais plutôt selon des paramètres de similitude acoustique entre les voix présentées. Ces paramètres sont bien connus et il existe un certain degré de standardisation dans la sélection de leurs valeurs. Par exemple, Hollien (1990), Eriksson (2007), Nolan (2003) et certains autres dressent une liste exhaustive de paramètres à contrôler (on examinera en détail ces critères à la section 1.1.).

Il est important de reconnaître que l'identification de la voix dans une technique de parade vocale ne repose pas sur l'hypothèse qu'il existe une « empreinte vocale » que l'on peut mesurer objectivement. Cette hypothèse, qui fait référence au fait que certains aspects acoustiques ont des caractéristiques spécifiques à l'individu, sous-tend que la production d'un même énoncé par un locuteur unique comporterait des attributs spectrographiques similaires, mais néanmoins spécifiques à l'individu (French 1994). Au début, des auteurs ont tenté d'appliquer cette idée d'empreinte vocale à l'analyse de spectrogrammes, mais, tel que mentionné par French (1994), certains exemples célèbres ont discrédité cette approche. Toutefois, cette même hypothèse sous-tend l'application de procédés de reconnaissance automatique. Comme on le verra dans la recension de la littérature ci-dessous (section 1.1.4), les techniques axées sur la perception humaine de voix familières offrent plusieurs avantages en comparaison aux techniques de reconnaissance automatiques.

Après des décennies de tentatives, on reconnait de façon générale les limites d'une approche automatisée. D'ailleurs, la reconnaissance automatique de l'empreinte vocale,

analogie de l'empreinte digitale, ne peut être considérée comme une preuve en cour, que ce soit en Amérique du Nord ou en Europe (voir, p.ex. Boe (2000); Eriksson (2007); Farrus (2009); French (1994); Hollien (1990)). À titre d'exemple de la difficulté à associer certaines variables acoustiques à des individus, la figure 1 illustre le cas de variations acoustiques dans la parole d'un même locuteur, lesquelles variations peuvent être très marquées (on voit p.ex., des différences dans les formants vocaliques, les durées de fermeture, la largeur des bandes de bruit, etc.). Il a été démontré qu'un manque de détection de telles variations par des techniques automatisées peut générer de fausses identifications positives. Comme le remarquent certains auteurs (Abo-Sahhad, Ahmed, et Abbas, 2014; Jain et coll., 2007; Unar et coll. (2014)), l'identification automatisée par la voix est le facteur biométrique ayant le moins de précision en comparaison avec celles effectuées à l'aide des empreintes digitales, du visage ou encore de l'iris.

Figure 1. *Deux spectrogrammes différents pour le même énoncé produit par le même locuteur avec différents accents régionaux. Tiré de* French, P. (1994).



Toutefois, par comparaison à des techniques objectives automatisées, les techniques faisant usage de la capacité humaine d'identification de la voix, bien que plus variables, présente la possibilité de fort taux d'identifications correctes dans la mesure où il est possible de contrôler certains facteurs individuels, surtout la « familiarité ». Comme l'indiquent Amino, Osanai, Kamada, Makinae, et Arai (2012b), la familiarité est un facteur clé dans l'identification correcte d'un locuteur. Cette notion sera abordée plus en profondeur à la section 1.1.2. Autrement dit, une parade vocale élaborée avec rigueur représente une technique d'identification par la voix plus efficace que les techniques automatisées.

Notons que les cas légaux pour lesquels la parade vocale est utilisée sont, en grande majorité, des cas pour lesquels le manque de preuve rend le jugement ardu. Leur utilisation peut donc avoir une importance non négligeable dans le verdict de culpabilité ou d'innocence d'un suspect. On peut penser au cas de menaces reçues par téléphone ou contenues sur des enregistrements qui, s'il y avait un moyen sûr d'identifier les individus, pourraient constituer des éléments de preuve qui ne seraient pas simplement circonstanciels. Bien que la technique des parades vocales ne soit utilisée qu'occasionnellement et puisse ainsi sembler marginale, son apport est parfois déterminant dans l'identification ou la discrimination d'un suspect. Plusieurs auteurs soulignent cette importance et invitent à une utilisation minutieuse et juste de la parade vocale ainsi qu'à un approfondissement des connaissances portant sur les implications que pourraient avoir certaines variables (Braun et Künzel, 1998; Broeders et van Amelsvoort, 1999; Butcher, 1996; Eriksson, 2007; Hollien, 1990; Hollien, Huntley Bahr, et Harnsberger, 2014; Hollien, Huntley Bahr, Künzel, et Hollien, 1995; Jessen, 2008; Künzel, 1994; Nolan, 2003; Philippon et coll., 2007b; Philippon, Cherryman, Vrij, et Bull, 2008; Wilding, Cook, et Davis, 2000; Yarmey, 1995, 2001). Parmi celles-ci, la familiarité entre l'individu chargé d'identifier une voix et celui qui parle est d'une importance centrale tel que le soulignent Yarmey, Yarmey, Yarmey, et Parliament (2001), Amino et coll. (2012b) et Amino, Osanai, Kamada, Makinae, et Arai (2012a). Certaines conditions d'écoute sont aussi essentielles à l'identification de voix familières. Par exemple, il semble fort pertinent d'observer à partir de quelle longueur de stimulus la familiarité devient un facteur clé (Bricker et Pruzansky, 1966; Hollien, 1990; Pollack, Pickett, et Sumby, 1954).

Par rapport à ces variables, la recension de la littérature aux sections suivantes permettra de faire le point sur certaines lacunes sur le plan des connaissances sur les effets systématiques de la familiarité dans les parades vocales. Cette recension permettra aussi de souligner les avantages qu'offre une technique utilisant la perception humaine par comparaison aux techniques automatisées. Ces sections, qui présentent, entre autres, un bilan des connaissances sur les effets de la familiarité, nous amèneront à formuler des hypothèses (section 2) et une expérience (section 3) portant sur certaines variables clés associées à la familiarité des voix et à la longueur des échantillons de parole dans l'identification vocale.

1.1 Recension de la littérature

Pour clarifier la question des effets de la familiarité dans l'identification de la voix en situation de parade vocale, il est important de voir, de prime abord, comment une parade vocale est élaborée. Les paragraphes qui suivent résument les propriétés d'une parade vocale qui font consensus et certaines considérations utiles portant sur la familiarité et les effets possibles de la longueur des stimuli sur la reconnaissance vocale. On verra aussi comment la technique de parade vocale se compare, souvent avantageusement, aux techniques automatisées de reconnaissance de la voix (ou reconnaissance de locuteurs).

1.1.1 Élaboration d'une parade vocale

Comme on l'a indiqué, l'identification d'un individu par la voix peut s'avérer cruciale dans le verdict d'un juge ou d'un jury. Cela a été illustré récemment dans le cas célèbre de Zimmerman en Floride en 2013. Dans ce cas, l'enregistrement d'un appel au 911 effectué par un résident du quartier où le drame a eu lieu comprenait des cris lors du crime. Cette preuve c'est avérée cruciale lors du procès pour déterminer lequel des deux individus impliqués, Georges Zimmerman ou Trayvon Martin, était l'agresseur. Après avoir affirmé que les cris provenant de l'enregistrement n'étaient pas scientifiquement identifiables, l'expert en reconnaissance de la voix du FBI, Hirotaka Nakasone, a soutenu que la procédure la plus fiable serait une reconnaissance de la voix par un individu familier :

> «*[The] best approach would be familiar voice recognition by an individual who has heard him ...speaking uttering in a variety of conditions including screaming, yelling under similar setups*.».(Donaghue, 2013)

Dans le cadre de la *phonétique forensique* (de l'angl. *forensic phonetics*), qui demeure un domaine de recherche relativement jeune, plusieurs auteurs se sont intéressés à l'élaboration de protocoles d'identification de locuteurs dans un cadre légal, ce qui a mené à l'utilisation de la méthode de «parade vocale» (par analogie à la parade visuelle de suspects; voir en particulier (Boe, 2000; Braun et Künzel, 1998; Broeders et van Amelsvoort, 1999; Butcher, 1996; Eriksson, 2007; French, 1994; Hollien, 1990; Hollien et coll., 2014; Hollien et coll., 1995; Interpol, 2001; Jessen, 2008; Künzel, 1994; Nolan, 2003; Philippon et coll., 2007b; Wilding et coll., 2000; Yarmey, 1995, 2001). Dans la littérature portant sur la parade vocale et l'identification de locuteurs, il existe un consensus sur certains critères à respecter dans l'élaboration des parades pour des fins légales. En fait, on reconnaît que pour avoir une identification valable, les voix dans une parade doivent refléter des locuteurs de même âge, de même sexe, ayant le même dialecte (ou accent) et une fréquence fondamentale moyenne parlée (F0mp) comparable; de plus, les échantillons de parole doivent présenter un débit similaire et refléter des énoncés de mêmes longueurs (Broeders et van Amelsvoort, 1999; Eriksson, 2007; Hollien, 1990; Hollien et coll., 2014; Jessen, 2008). La qualité des enregistrements est elle aussi fréquemment soulignée dans la littérature (voir p. ex. Betancourt et Huntley Bahr (2010)). Un manque de rigueur dans le contrôle de ces multiples facteurs peut résulter en un biais important sur le plan de la discrimination des individus dans une parade.

Plus précisément, il a été rapporté que l'altération de certains paramètres acoustiques liés au vieillissement peut être perceptible et par conséquent affecter la tâche d'identification en parade vocale (Braun, 1996; Dilley, Wieland, Gamache, Devin McAuley, et Redford, 2013). En particulier, un enfant est facilement différenciable d'un adulte, tout comme une personne âgée serait difficilement confondue avec un jeune adulte. Dans son étude sur la perception de l'âge par différents groupes de participants, Braun (1996) fait état des corrélations entre l'âge réel et l'âge perçu des locuteurs observées par plusieurs auteurs. La méthodologie utilisée pour des études telles que celle de Braun (1996) est généralement très simple : un nombre de participants reçoit comme directive d'écouter plusieurs enregistrements de locuteurs et d'estimer leur âge.

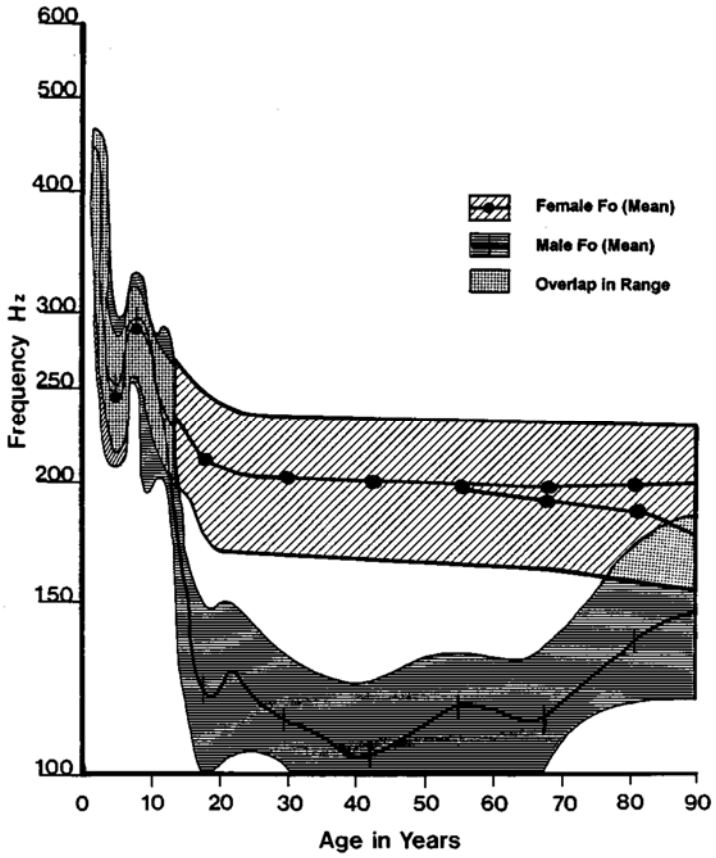
Tableau I. *Corrélations entre l'âge réel et l'âge perçu telles qu'établies par différents auteurs*. *Adapté de* Braun, A. (1996).

Auteurs (date)	Corrélation
Shipp et Hollien (1969)	0.88
Ryan et Burk (1974)	0.77
Nieman et Applegate (1990)	0.88
Horii et Ryan (1981)	0.76
Ramig, Scherer, et Titze (1985)	0.17
Braun (1996) experts	0.70
Braun (1996) étudiants	0.68

Les résultats du Tableau I illustrent, à l'aide de plusieurs coefficients de corrélation positifs, qu'il est important de s'assurer que les individus proviennent d'un même groupe d'âge et possèdent des attributs vocaux similaires lors de la sélection des voix d'une parade vocale.

En particulier, une des variables incontournables dans la constitution de la parade porte sur la fréquence fondamentale (F0) et la F0mp. Il est bien établi que la F0 varie selon le sexe. Bien que certains individus possèdent une voix partageant certaines caractéristiques typiquement associées au sexe opposé, l'utilisation de voix de même genre dans l'élaboration d'une parade vocale est reconnue comme fondamentale (Hollien, Hollien, et de Jong, 1997; Liu, Chen, Jones, Wang, Chen, Huang, et Liu, 2013; Sandmann, am Zehnhoff-Diennesen, Claus-Michael, Rosslau, Lang-Roth, Burgmer, Knief, Matulat, Vauth, et Deuster, 2014). La figure 2 illustre les différences de F0mp entre les hommes et les femmes à différents âges telles que rapportées par Hollien et coll. (1997). Comme on peut l'observer, cette différence est marquée à tous les âges entre 15 ans et 75 ans.

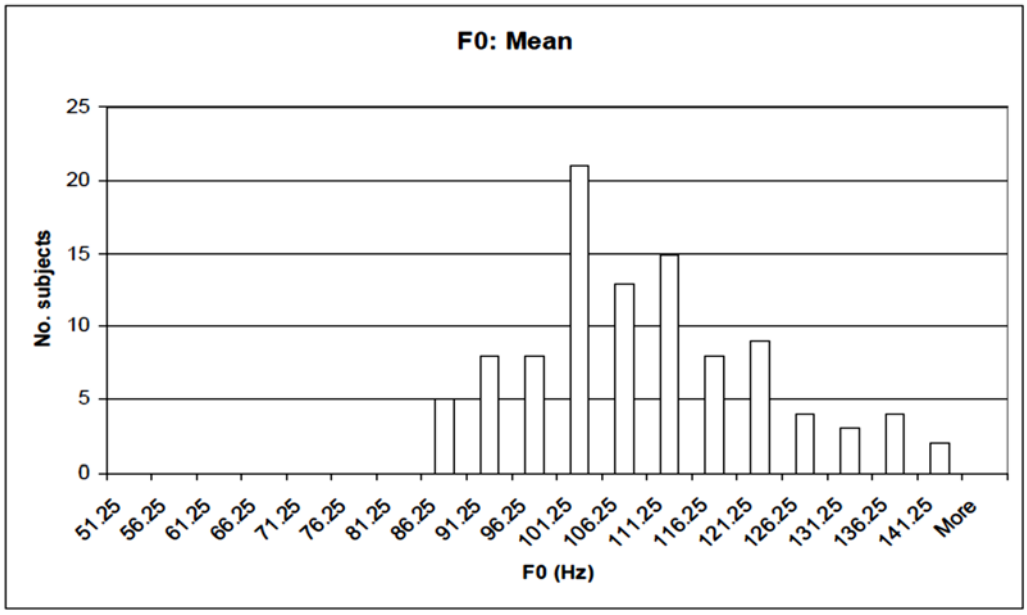
Figure 2. *Généralisation des F0mp pour les hommes et les femmes en function de l'âge. Tiré de* Hollien, H., Hollien, P., & de Jong, G. (1997).



Plusieurs autres auteurs notent le fait que la F0mp doit être contrôlée dans l'élaboration d'une parade vocale même lorsque le sexe des locuteurs est préalablement contrôlé (Eriksson, 2007; Foulkes et Barron, 2000; French, 1994; Hudson, de Jong, McDougall, Harrison, et Nolan, 2007; Jessen, 2008; Kinoshita, Ishihara, et Rose, 2009). Par exemple, Foulkes et Barron (2000) ont effectué des analyses de F0mp sur les voix des participants à leur étude afin d'expliquer pourquoi certaines d'entre elles étaient mieux identifiées que d'autres. Leurs analyses ont démontré que les voix ayant des valeurs plus extrêmes de F0mp sont plus facilement identifiables en comparaison avec des voix dont la F0mp se situe plus près de la moyenne. La figure 3 démontre quant à elle le large écart de F0mp observable chez des individus provenant d'une population relativement restreinte telle qu'observée par Hudson et coll. (2007). On comprend donc que, même une fois l'âge, le genre et le dialecte contrôlés, la

F0mp des locuteurs est un élément de variation acoustique important pouvant biaiser l'identification dans le cadre d'une parade vocale.

Figure 3. *F0 moyenne en discours spontané de 100 jeunes hommes âgés entre 18 et 25 ans locuteurs natifs de l'anglais standard du sud de la Grande-Bretagne. Tiré de* Hudson et coll. (2007).



Par ailleurs, Kinoshita et coll. (2009) ont tenté de démontrer que la F0mp et la variance de la F0 ne sont pas des indices acoustiques suffisants pour une identification vocale fiable. Ces auteurs ont utilisé d'autres indices reliés à la F0 dont un coefficient de dissymétrie (angl. *skewness*), un coefficient d'aplatissement (angl. *kurtosis*), la F0 modale et la densité de la F0mp. Leurs résultats ont néanmoins démontré que la F0mp et ses écarts-types étaient, même lorsqu'utilisés indépendamment des autres indices mesurés, pertinents lors d'une identification par la voix. On peut en conclure qu'une variation plus ou moins importante de la F0mp chez les différents individus dont la voix est présente dans une parade affecte directement le niveau de difficulté associé à la tâche d'identification

Une différence de dialecte peut, elle aussi, biaiser l'identification dans une situation de parade vocale. Scharinger, Manahan et Idsardi (2013) ont démontré par le biais de la technique de potentiels évoqués que l'être humain perçoit très rapidement (< 200 ms.) une variation

dialectale dans la parole, et ce, lors de la présentation d'énoncés d'une durée aussi courte que deux syllabes. Pour se faire, les expérimentateurs ont présenté plusieurs occurrences d'un énoncé d'une longueur de deux syllabes (durée moyenne de 500 ms.) prononcées en anglais américain standard et en anglais afro-américain vernaculaire à 15 participants. Ces participants étaient tous locuteurs natifs de l'anglais américain standard et avaient un niveau d'éducation similaire. Aucune tâche ne leur était attribuée, autre que d'écouter les enregistrements qui leur étaient présentés. Leurs résultats démontrent une variation nette de l'amplitude de la MMN, supportant une perception très rapide d'une variation dialectale par l'être humain. L'importance de prendre en compte cette variation dialectale est généralement reconnue dans la littérature puisque certains paramètres acoustiques y étant liés peuvent facilement être perceptibles par un identificateur (Broeders et van Amelsvoort, 1999; Eriksson, 2007; Foulkes et Barron, 2000; Kerstholt, Jansen, van Amelsvoort, et Broeders, 2006; Scharinger et coll., 2011). De plus, la langue et le dialecte parlés par l'identificateur lui-même doivent préférablement être les mêmes que ceux du locuteur à identifier de façon à éviter une procédure biaisée (Philippon, Cherryman, Bull, et Vrij, 2007a).

Finalement, la qualité des enregistrements servant à l'identification lors d'une parade vocale a fait l'objet de plusieurs études et peut avoir un effet déterminant sur les taux d'identification. Il a été établi que les effets de la qualité acoustique associée à une communication directe, téléphonique ou cellulaire peuvent nuire à l'identification des voix de locuteurs (Betancourt et Huntley Bahr, 2010; Broeders et van Amelsvoort, 1999; Foulkes et Barron, 2000; Harrison, 2001; Kerstholt et coll., 2006; Künzel, 2001; Meinerz et Masthoff, 2011; Nolan, 2002). En particulier, Betancourt et Huntley Bahr (2010) ont examiné les effets de la variation de langue, de dialecte et de qualité d'enregistrement lors de tâches d'identification. Dans les expériences qu'elles ont menées, les participants entendaient des paires d'énoncés qui variaient selon 3 variables : la qualité de l'enregistrement, la langue parlée et la modalité de production (énoncé lu ou prononcé en discours spontané). Suite à l'écoute d'une paire d'énoncés, les participants devaient indiquer le plus rapidement possible s'il était question d'un même locuteur ou de deux locuteurs différents. Leurs résultats démontrent que la variation de la qualité de l'enregistrement engendre une différence entre les taux d'identification plus importante que lorsque la langue ou le dialecte varient. De leur côté,

Kerstholt et coll. (2006) ont mené une étude afin de déterminer les effets d'une variation de l'accent, de l'utilisation du téléphone (en comparaison avec des enregistrements directs) et du temps écoulé entre l'exposition à la voix et la tâche d'identification. Encore une fois, plusieurs conditions représentant un changement par rapport aux variables mentionnées précédemment ont été utilisées. Contrairement à l'étude de Betancourt et Huntley Bahr (2010), les résultats de Kerstholt et coll. (2006) n'ont montré une variation des taux d'identification que par rapport à un changement dialectal, et aucune variation significative n'a été observée entre les enregistrements directs et ceux faits par téléphone. En somme, bien que les résultats puissent varier selon les études, la majorité des cas légaux pour lesquels la parade vocale est pertinente impliquent des crimes (menace, fraude, etc.) ayant été commis à l'aide d'un téléphone cellulaire, il apparaît nécessaire de reproduire ces conditions dans le but de mieux généraliser les résultats obtenus à des situations légales usuelles.

Pour résumer, les résultats des différentes études citées et décrites ci-dessus soulignent l'importance de plusieurs variables lors de l'élaboration de parades vocales. On peut en conclure qu'un manque de contrôle de l'âge, du sexe, des caractéristiques articulatoires (p. ex. F0mp et débit) et de l'accent (ou du dialecte) des voix constituant la parade vocale ainsi que de la qualité des enregistrements peut résulter en une discrimination facilitée pour l'identificateur.

1.1.2 La familiarité en parade vocale

Dans certains cas, le recours à la parade vocale implique des individus préalablement familiers. En soi, la familiarité d'une voix peut se définir en fonction du degré de contact entre des locuteurs. Bien que certaines études aient abordé le rôle potentiel de l'effet de familiarité des voix en contexte de parade vocale, il est surprenant de noter que celles-ci n'ont pas délimité de manière objective les différentes catégories de familiarité en fonction de paramètres quantifiables (Amino et coll., 2012a; Eriksson, 2007; Foulkes et Barron, 2000; Hollien, 1990; Hollien et Schwartz, 2000; Yarmey et coll., 2001). En fait, ces études sont fondées sur des catégories de familiarité définies selon des standards sociaux et culturels et ne permettent pas une quantification opérationnelle de la relation ou du degré de contact entre les individus concernés.

Par exemple, Yarmey et coll. (2001) se sont intéressés à la corrélation entre la performance estimée par les participants lors d'une tâche d'identification de voix familière et les résultats réellement obtenus. Les observations ont été faites dans une première condition pour laquelle les énoncés présentés étaient prononcés avec une intonation naturelle. La seconde condition impliquait des énoncés chuchotés. En ce qui a trait à la présentation des stimuli, la méthode utilisée était fortement inspirée de la parade vocale. Les participants devaient donc écouter un ensemble de voix dans lequel une voix familière était présente. Suite à l'écoute, ils devaient identifier la voix de l'individu familier. Il est important de souligner que la procédure expérimentale utilisée ne spécifiait que vaguement la nature de la familiarité entre l'identificateur et les voix à identifier.

> *«A high-familiar speaker was defined as an immediate family member or best friend of a particular listener; a moderate-familiar speaker was a co-worker, team mate, club-mate, or general friend of the listener; and a low-familiar speaker was defined a casual acquaintance, such as a next-door neighbour or associate, who would be expected to have talked with the listener for only a few minutes in any week over the last year. » Yarmey et coll. (2001) p. 287*

Tel que mentionné en introduction, cette classification en niveaux de familiarité semble fondée sur un jugement en fonction de standards sociaux et par conséquent très difficilement quantifiables. Néanmoins, les résultats qu'ils ont obtenus démontrent que, de manière générale, les taux d'identification obtenus par les participants étaient significativement plus faibles que ce que ceux-ci avaient estimé. De tels résultats démontrent la difficulté réelle associée à une tâche qui semble a priori aisée et l'importance de procéder à une identification par la voix selon une méthodologie fiable et non fondée sur l'idée reçue qu'une telle identification est possible même lorsque la familiarité est faible ou moyennement élevée. Ces résultats illustrent aussi le besoin réel de procéder à une meilleure quantification de la familiarité lorsque celle-ci est utilisée dans un contexte d'identification par la voix

De leur côté, Foulkes et Barron (2000) n'ont utilisé que des participants issus d'un même réseau social à familiarité dite élevée, c'est-à-dire que tous les participants possédaient sensiblement le même niveau de familiarité entre eux et qu'aucune variation de cette variable n'était observée. La familiarité entre les membres de ce réseau est ainsi décrite puisqu'ils avaient tous cohabité en résidence étudiante pendant la même période d'un an, soit l'année

précédant l'expérimentation. Les stimuli utilisés étaient des enregistrements des voix des membres de ce réseau social ainsi que de deux voix inconnues prononçant des énoncés de 8 à 10 secondes. Bien que ces énoncés étaient lus et prédéfinis, les participants ont reçu la directive de les prononcer aussi naturellement que possible. Les participants, toujours les membres de ce réseau, devaient ensuite écouter tous les enregistrements et identifier qui en était l'auteur en incluant la possibilité que la voix présentée soit la leur ou celle d'un inconnu. Suivant une telle méthodologie, il est impossible d'établir avec précision les paramètres constituant un niveau de familiarité élevé, moyen ou faible puisqu'aucun autre niveau de familiarité n'est observé ou décrit. Encore une fois, la procédure de présentation des stimuli était apparentée à celle d'une parade vocale, bien que la sélection des voix qui y apparaissant ne respectait pas les lignes directrices habituellement reconnues dans l'élaboration d'une parade vocale, en raison du design expérimental et des objectifs propres à l'étude. La grande variabilité présente dans les résultats obtenus ainsi que les analyses de F0 effectuées par les auteurs supportent l'importance du dialecte et de la F0mp lors d'une tâche d'identification par la voix (tel que soutenu à la section 1.1.1) puisque les voix les plus facilement reconnues étaient celles de locuteurs ayant un accent régional plus prononcé ou encore une F0mp significativement plus élevée ou basse que la moyenne des participants. Finalement, bien que la familiarité entre les participants était dite élevée, les écarts observés dans les taux d'identification permettent encore une fois de souligner l'importance d'une familiarité mieux définie.

Une étude semblable au niveau des résultats a été menée par Amino et coll. (2012a) pour laquelle les participants, les stimuli et la tâche expérimentale de étaient très similaire à ceux de l'étude de Foulkes et Barron (2000). Les résultats obtenus démontrent une claire amélioration des taux d'identification lorsque le locuteur est familier du participant. Les auteurs appuient donc l'approche selon laquelle la reconnaissance d'une voix familière n'implique par les mêmes processus cognitifs que lorsque la voix est connue sans pour autant être familière.

Finalement, dans son ouvrage référence, Hollien (1990) fait état des recherches effectuées en lien avec l'identification par la voix (Bricker et Pruzansky, 1966; McGehee, 1937, 1944) et soutient que les résultats de ces études varient significativement selon la longueur des énoncés présentés, variable faisant l'objet de discussion à la section 1.1.3, et non

selon la familiarité. L'étude de Hollien et Schwartz (2000) ayant comme objectif de déterminer l'effet de la récence des enregistrements sur l'identification par la voix apporte cependant de nouvelles conclusions sur la pertinence de la familiarité dans certains contextes. Dans cette étude, les participants devaient écouter des paires de deux enregistrements effectués à deux différents moments (allant d'immédiatement avant l'expérimentation à 20 ans plus tôt) et indiquer s'il s'agissait du même locuteur. Plusieurs variables, telles que le sexe des participants et leur formation en phonétique, ont été contrôlées et leurs effets vérifiés. L'une d'entre elles était la familiarité. Plusieurs participants étaient familiers avec une ou plusieurs des voix utilisées comme stimuli lors du premier enregistrement (20 ans auparavant), mais avaient perdu contact avec l'individu depuis. Malgré cette familiarité, aucun d'entre eux n'a pu associer correctement les voix comme provenant d'un même locuteur. Les auteurs concluent donc que, dans le cadre de leur expérience, la familiarité n'a eu aucun effet. Il importe de souligner que les auteurs mentionnent cette familiarité sans la définir en termes quantifiables, ce qui peut porter à croire que la familiarité, suite à une longue période de temps, n'a aucun effet.

L'importance de l'effet de la familiarité, lorsqu'on considère la capacité de l'humain à reconnaître un parent ou un proche seulement à l'aide de sa voix, suggère que l'amélioration de l'identification vocale nécessite une attention particulière sur la façon dont on mesure la familiarité des voix. Les différentes études décrites ci-dessus, bien que portant une attention particulière à la familiarité des voix, ne quantifient pas cette notion et arrivent, par conséquent, à des conclusions qui pourraient être remises en question. L'apport de la présente étude est donc de rendre cette variable opérationnelle en termes de certains critères mesurables, soit la *récence*, la *fréquence*, la *durée* et la *période* des contacts (voir la section 3)

1.1.3 La longueur des énoncés présentés dans les parades vocales

Mise à part la variable centrale de la familiarité, il est reconnu que la longueur des stimuli acoustiques présentés lors d'une tâche d'identification a un effet sur le taux d'identification. Plusieurs travaux portant sur les procédures des parades vocales soutiennent que plus le stimulus est long, meilleure sera l'identification des locuteurs, mais il n'y a pas de consensus sur une longueur optimale (Amino et coll., 2012b; Bricker et Pruzansky, 1966; Broeders et van Amelsvoort, 1999; Hollien, 1990; Pollack et coll., 1954). Les données

apportées par Bricker et Pruzansky (1966) et Pollack et coll. (1954) soutiennent cependant l'hypothèse que le contenu du stimulus en terme d'inventaire de sons produits est plus pertinent que la durée elle-même du stimulus.

Les résultats de Pollack et coll. (1954)[1] par exemple montrent que la durée n'est qu'indirectement associée à une amélioration des taux d'identification et que la taille de l'échantillon des différentes productions articulatoires est le principal facteur. Aussi, leurs résultats indiquent que l'amélioration de l'identification due à la durée sature à partir de 1.2 seconde. Les auteurs soulignent cependant que la quantité d'information nécessaire à une identification correcte augmente en fonction du nombre de voix parmi lesquelles la voix à identifier se trouve. Notons que cette étude impliquait l'identification de voix familières.

Pour ce qui est de l'étude de Bricker et Pruzansky (1966), ces auteurs ont tenté de vérifier si l'identification vocale d'un individu variait en fonction du nombre de phonèmes présent dans un stimulus. Dans leur étude, les 16 participants ayant pris part à une tâche d'identification étaient familiers avec les 10 voix utilisées comme stimuli avant l'expérimentation. Une tâche de familiarisation comprenant une confirmation de l'identité du locuteur entendu a été utilisée au début de la tâche afin de minimiser les effets de la variation de familiarité entre les participants et les voix utilisées comme stimuli. Ces stimuli étaient des enregistrements de monosyllabes, de dissyllabes, d'énoncés, de voyelles extraites des énoncés et de suites consonne-voyelle (CV) aussi extraites des énoncés. Pour la tâche expérimentale, qui s'étendait sur trois jours, les participants n'avaient qu'à écouter les stimuli et indiquer, sur un formulaire, le prénom du locuteur présumé. Le Tableau II résume les résultats de l'étude et démontre une augmentation du taux d'identification en fonction du nombre de phonèmes présentés.

[1] Bien que la pertinence des résultats exposés dans cette étude soit explicite et fortement reconnue dans la littérature, la méthodologie utilisée par Pollack et coll. (1954) est opaque et parfois laconique. L'étude en question demeure néanmoins l'une des plus citées et pertinente à l'égard des la longueur des stimuli utilisés en tâche d'identification par la voix.

Tableau II. *Taux d'identification, durée et nombre de phonèmes pour les 5 types de stimuli utilisés. Adapté de* Bricker et Pruzansky (1966).

	% correct	Nb. de phonèmes	Durée (ms.) Moyenne	É.t.
Énoncés	98	>15	2400	500
Dissyllabes	87	4	446	83
Monosyllabes	81	3.2	498	118
Suites CV	63	4	117	15
Voyelles	56	1	117	15

Globalement, ces résultats indiquent qu'une hausse des taux d'identification en fonction du nombre de phonèmes présentés est observable, mais que cet effet tend à plafonner au-delà de 4 phonèmes.

Globalement, les résultats des études de Pollack et coll. (1954) et Bricker et Pruzansky (1966) illustrent que l'augmentation de la durée des énoncés présentés lors d'une identification par la voix n'est pas en soi responsable d'une amélioration de l'identification. En fait, leurs résultats démontrent que la taille de l'échantillon en nombres d'éléments transcrits (« phonèmes » encore configuration articulatoire) plutôt que la durée des stimuli le facteur ayant le plus d'impact sur une hausse des taux d'identification.

1.1.4 Les avantages de l'identification vocale humaine par rapport aux techniques automatisées

Pour plusieurs auteurs, les techniques automatisées sont à privilégier par rapport aux techniques axées sur la perception humaine (Gonzalez-Rodriguez, Torre Toledano, et Ortega-Garcia, 2007; Wang, Chen, et Chi, 2002). Ce jugement est cependant moins qu'évident à l'heure actuelle et il importe de considérer les avantages et inconvénients de ces différentes approches dans l'état présent des techniques automatisées. Le tableau III ci-dessous résume par points de comparaison différents avantages et inconvénients des méthodes qui peuvent faire l'objet de polémiques.

Tableau III. *Comparaison entre les avantages et les inconvénients de la perception humaine en parade vocale et des techniques automatisées d'identification par la voix.*

	Paramètres	Parades vocales	Techniques automatisées
1	Aucun entrainement[1]	✓	_
2	Conditions généralisables à un contexte légal[2, 3]	✓	_
3	Considère l'information contextuelle[4]	✓	_
4	Coût matériel moindre	✓	_

1 McLaren, Lawson, Ferrer, Scheffer, et Lei (2014)

2 Adibi (2014)

3 Al-Hudhud, Abdulaziz Alzamel, Alattas, et Alwabil (2014)

4 Jessen (2008)

Il est important de souligner le premier point au tableau III qui porte sur les limites générales des techniques automatisées. Toutes ces techniques nécessitent une période d'entraînement dont la durée est indéterminée. Lorsque de nouvelles conditions d'enregistrement sont rencontrées, une nouvelle période d'entraînement est requise et les logiciels doivent être adaptés (McLaren et coll., 2014). Par contre, l'être humain est apte à identifier plusieurs voix provenant d'une grande variété d'environnements sans aucune adaptation. Notons aussi que, comme le mentionnent Gonzalez-Rodriguez et coll. (2007) et Unar et coll. (2014), les techniques d'identification automatisées sont plus fréquemment utilisées dans un contexte commercial et l'identification est souvent appuyée par des informations secondaires (p. ex. un mot de passe verbal) afin de confirmer l'identité de l'individu. En comparaison, l'identification par un humain ne nécessite aucun entraînement puisqu'un identificateur possède déjà l'information nécessaire en mémoire.

Mentionnons, en ce qui a trait au point 2 du tableau III, que les résultats d'une forte majorité d'études portant sur l'identification automatique par la voix sont obtenus dans des conditions d'enregistrement optimales et hautement contrôlées (p. ex. Adibi (2014) et Al-Hudhud et coll. (2014)). Ces résultats sont par conséquent peu applicables aux conditions bruyantes fréquemment rencontrées lors de communications litigieuses par téléphones (ex. menaces, fraudes, etc.) et autres crimes n'ayant comme preuve que des données acoustiques.

Par exemple, il serait improbable d'obtenir un enregistrement potentiellement incriminant fait depuis une salle insonorisée et respectant les normes d'enregistrement admises dans le milieu scientifique.

Pour ce qui est de l'information contextuelle, au point 3 du tableau III, Jessen (2008) fait état de la situation actuelle en ce qui concerne les différentes approches ainsi que les méthodologies généralement adoptées dans le domaine de la phonétique légale. En concluant sa section sur l'identification de locuteurs, il écrit :

> *«Automatic speaker identification, for example, scans the speech signal more comprehensively than analytical methods, so that hardly any potential speaker specific component will be left out, yet on the other hand, through the expert knowledge that is required in analytical methods, those pieces of information can be identified that are of particular importance for the speaker identification process and which might get lost in the wealth of information collected with automatic methods. For example, presently automatic systems would not usually be able to detect the type of dialectal information that leads to the conclusion that two speakers are probably not identical».(Jessen, 2008)*

L'auteur souligne ici que les méthodes automatisées ne sont pas en mesure de considérer un nombre important d'informations acoustiques, telles que le dialecte parlé, pourtant nécessaires afin de procéder à une identification fiable. Il soutient aussi que lorsque l'analyse est effectuée par un expert, en l'occurrence lors de parades vocales, ces informations sont alors prises en considération. Les observations plus récentes faites par Amino et coll. (2012b) suggèrent que les techniques automatisées d'identification par la voix font toujours face à ces problèmes liés aux informations contextuelles contenues dans la parole.

Enfin, par rapport aux taux de succès des techniques automatisées, Jain et coll. (2007) et Unar et coll. (2014) font état d'un taux de fausses identifications positives (2-5 %) ainsi que d'un taux de fausses discriminations (5-10 %) pour les techniques automatisées d'identification par la voix. Notons que lorsqu'il est question de situations légales pouvant entraîner des verdicts biaisés, ces marges d'erreur sont très grandes. Des études plus récentes (p. ex. Kanagsundaram et coll. 2014) tentent de pallier à la situation, mais les résultats obtenus font toujours preuve d'une marge d'erreur trop élevée pour permettre une application dans la sphère légale.

En somme, l'utilisation de techniques basées sur la perception humaine de la voix dans l'identification de locuteurs en contexte légal possède plusieurs avantages sur les techniques automatisées. Ce sont ces multiples avantages qui motivent la présente recherche où une opérationnalisation du facteur de la familiarité est vue comme un facteur clé de l'identification vocale efficace.

2 Objectifs de l'étude et hypothèses de recherche

La section précédente met en lumière les avantages d'une technique de reconnaissance des voix qui fait usage de la perception humaine plutôt que des procédés automatiques. Cependant, l'exactitude de la reconnaissance humaine repose sur un critère de familiarité : plus une voix est familière, plus on accroît le taux de reconnaissance. Suivant ce simple principe, la présente étude vise à établir les conditions pouvant maximiser la reconnaissance humaine des voix avec une attention particulière portée à la quantification de la familiarité comme condition essentielle et à la longueur des énoncés présentés.

2.1 Objectif 1

Le premier objectif de l'expérience est d'observer l'effet de la familiarité d'une voix sur le taux de reconnaissance de locuteurs et d'établir un index opérationnel de familiarité axé sur quatre mesures du degré de contact entre individus, soit la *fréquence*, la *récence*, la *durée* et la *période* du contact. Cet index est établi dans le but de pallier à un manque d'opérationnalité du concept de familiarité dans la littérature (Amino et coll., 2012a; Foulkes et Barron, 2000; Hollien et Schwartz, 2000; Yarmey et coll., 2001)

2.2 Hypothèse 1

Suivant notre index de familiarité, on avance l'hypothèse que, plus la familiarité d'une voix est élevée dans une parade vocale, plus le taux de reconnaissance vocale du locuteur est élevé. Par conséquent, un taux d'identification correcte de 100 % est possible lorsque la familiarité est suffisamment élevée.

2.3 Objectif 2

Le second objectif est d'observer, dans un contexte de parade vocale, si la longueur des énoncés présentés a un effet significatif sur l'identification des voix. Une vérification de cet effet tient compte de résultats antérieurs démontrant que le nombre de phonèmes présents dans le stimulus influence davantage le taux d'identification que la durée du stimulus (Bricker et Pruzansky, 1966; Pollack et coll., 1954).

2.4 Hypothèse 2

En appliquant les critères conventionnels d'une parade vocale, on examinera l'hypothèse selon laquelle l'augmentation de la durée des énoncés (et de fait du nombre de phonèmes ou de configurations articulatoires) améliore le taux d'identification, mais que cet effet présente des limites.

3 Méthodologie

Une méthodologie particulière a été utilisée pour élaborer des parades vocales où on devait recruter des participants selon leur degré de familiarité avec une voix. Par exemple, après avoir enregistré la voix d'un locuteur, on devait recruter des individus pour qui ce même locuteur était plus ou moins familier (un frère, un ami, une connaissance, etc.). Avec ce type de démarche, on ne peut appliquer un recrutement quasi aléatoire quant aux caractéristiques des sujets (comme c'est généralement le cas pour les designs expérimentaux). Cependant, notons que tous les participants aux tâches décrites ci-dessous étaient âgés de plus de 18 ans. Par ailleurs, aucun des participants ne présentait de trouble de la parole ou d'audition apparent ou diagnostiqué et ces individus n'étaient pas sous médication lors des expériences. Mentionnons aussi que le Comité d'éthique de la recherche en arts et en sciences (CÉRAS) de l'Université de Montréal a approuvé, au préalable, la méthodologie ainsi que les procédures de recrutement adoptées.

3.1 Participants et questionnaire de familiarité

Quarante-quatre individus (26 femmes et 18 hommes) âgés de 18 à 60 ans (moyenne de 33 ans) ont participé à la présente étude. Tous les sujets avaient le français québécois comme langue maternelle et aucun ne possédait d'accent régional marqué. Les participants étaient répartis en trois groupes (deux groupes de 16 participants et un groupe de 12 participants); chaque groupe étant jumelé avec une « voix-cible » d'un individu dans une parade vocale (décrite à la section suivante). Ce jumelage des participants à des parades vocales se faisait selon quatre niveaux de familiarité entre le participant et l'individu avec la voix-cible. Ces quatre niveaux étaient établis au moyen d'un questionnaire pondéré[2] qui permettait de quantifier la communication entre les individus selon les variables suivantes (voir aussi la section 4) :

i. *La récence* des communications; à quand remonte la dernière conversation?
ii. *La fréquence* moyenne des communications; combien de fois par jour/année, en moyenne, ont lieu les conversations?

[2] Pour des raisons de propriété intellectuelle, le système de pointage et le barème qui y est associé sont confidentiels.

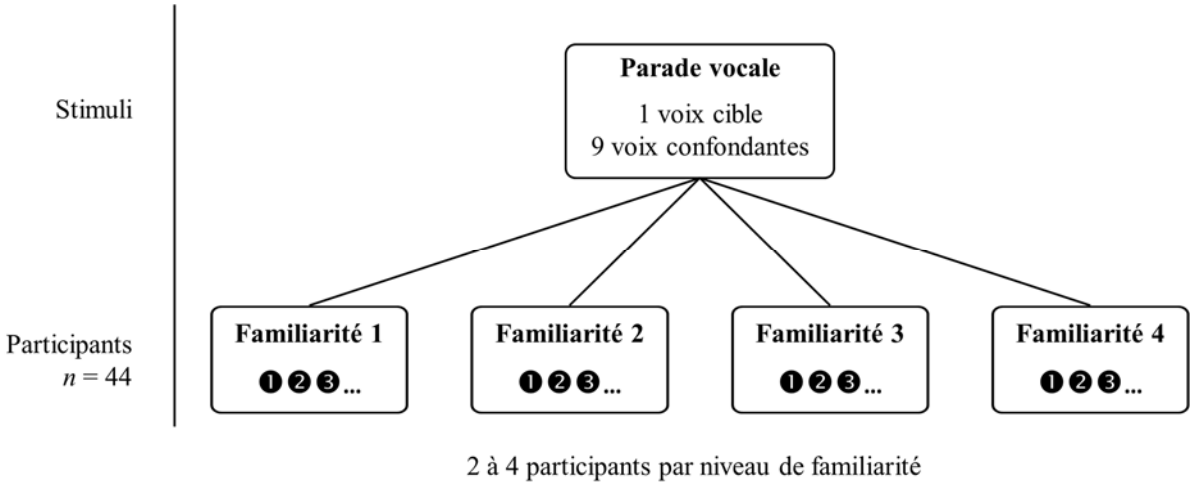
iii. *La durée* moyenne des communications; combien de temps, en moyenne, durent-elles?
iv. *La période* des communications; pendant combien de temps ont-elles eu lieu; depuis quand ont-elles lieu?

Les réponses à chacun des critères ci-dessus étaient pondérées. Par exemple, une réponse considérée comme étant à valeur faible pour l'un ou l'autre des quatre critères donnait 2.5 points; une réponse à valeur moyenne donnait 5 points; une réponse à valeur moyenne/élevée donnait 7.5 points et une réponse à valeur très élevée donnait 10 points. Un participant avec un index de familiarité 1 avait en moyenne un score de 4.38/10, tel qu'évalué par le questionnaire; un participant avec un index de familiarité 2 avait, quant à lui, un score moyen de 7.44/10; pour un index de niveau 3, le score moyen était de 8.31/10; et pour les participants de l'index le plus élevé (4), le score moyen était de 7.84/10.

Mentionnons que les quatre critères ci-dessus ont été utilisés parce qu'ils permettent de quantifier de façon efficace et objective les contacts entre locuteurs. En fait, sans avoir été définis comme « critères de familiarité », certaines des variables ci-dessus sont déjà reconnus et ont fait l'objet d'études. Par exemple, la méthodologie de l'étude de Hollien et Schwartz (2000) a permis d'observer les effets de la récence des communications (voir section 1.1.2).

Pour illustrer le regroupement des participants selon les critères de familiarité, la figure 4 présente un exemple d'une parade vocale où le stimulus comprend dix voix dont une voix-cible d'un individu qui est plus ou moins connu des participants. Le questionnaire permettait de classifier ce degré de connaissance par rapport au niveau de communication entre le participant et l'individu avec la voix-cible. Dans la présente étude, il y avait trois regroupements suivant la structure à la figure 4.

Figure 4. *Exemple d'une parade vocale et de la répartition des participants qui y sont associés. La présente étude utilise trois parades impliquant 44 participants au total.*



3.2 Les stimuli : élaboration des parades vocales

Tel qu'illustré à la figure 4, trois parades vocales qui contenaient chacune 10 voix (une voix-cible, neuf voix confondantes) ont été élaborées. Les procédures d'enregistrement et de recrutement utilisées étaient identiques pour les trois parades.

Les locuteurs présentant les trois voix-cibles enregistrées ont été recrutés selon deux critères clés : il fallait que les trois locuteurs consentent à ce que l'expérimentateur puisse avoir accès à leurs réseaux sociaux (famille, amis, connaissances, collègues, etc.) et ils devaient avoir le français québécois comme langue maternelle, sans accent régional prononcé.

Les locuteurs devaient produire, avec une voix neutre dans une salle à l'épreuve du bruit, huit énoncés spécifiques et un neuvième énoncé depuis un téléphone cellulaire (marque *LG* modèle 3200). Les enregistrements effectués depuis la salle à l'épreuve du bruit ont été faits à l'aide d'un microphone (marque *Electro Voice* modèle A635) relié à un ordinateur portable muni d'une carte de son de 32 bits. Les enregistrements effectués depuis le téléphone cellulaire étaient quant à eux transmis à un téléphone linaire (*Cisco* modèle IP Phone 7962) relié à un PC *(HP Compaq* modèle 6200 Pro avec carte de son 32 bits) au moyen d'un adaptateur (marque *Konexx* modèle 100). Tous les enregistrements ont été faits à l'aide du

même logiciel (*Goldwave*, v5.66) à un taux d'échantillonnage de 44,1 kilohertz (kHz) et encodés en format *.wav.*

Le tableau IV ci-dessous présente différents énoncés qui ont constitué les stimuli de l'expérience d'identification et leurs caractéristiques. Notons que le contenu des énoncés reflète la parole usuelle et qu'on ne peut identifier un individu sur la base du contenu exprimé. Par ailleurs, ces énoncés variaient en longueur et ils se divisaient en terme du degré de nasalité des sons produits puisque, tel que le mentionnent Pollack et coll. (1954) et Bricker et Pruzansky (1966), plus un énoncé est long (plus son « inventaire phonémique » est grand), plus les chances d'identification correcte sont fortes. Aussi, la présence de sons nasaux faciliterait l'identification d'un individu par sa voix puisque les cavités nasales ne sont pas altérées par le mouvement d'articulateurs comme c'est le cas pour la cavité orale (Amino et coll., 2012a; Glenn et Kleiner, 1967; Su, Li, et Fu, 1974). Les sons étant produits à l'aide des cavités nasales posséderaient donc une résonnance plus spécifique à l'individu que ceux produits uniquement par la cavité orale.

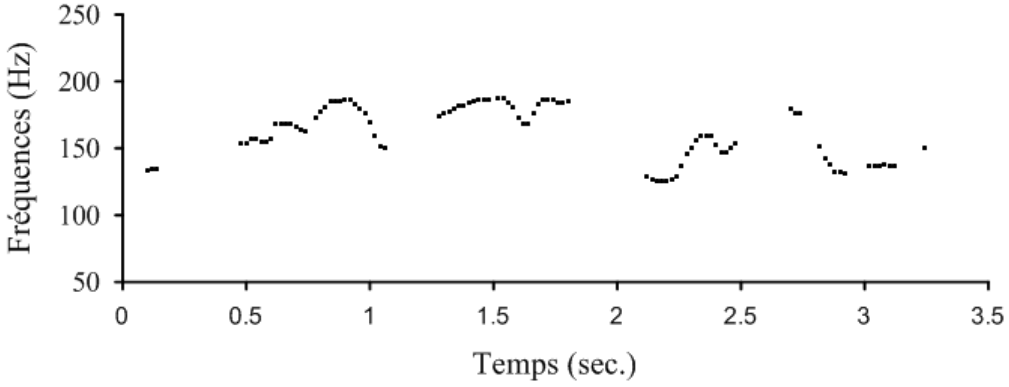
Tableau IV. *Exemples de stimuli utilisés selon la longueur en syllabe et l'utilisation ou non de cavités nasales.* P_{nasal} *= nombre de symboles API reflétant des éléments produits avec des résonnances nasales*. P_{total} *= nombre de symboles API pour l'énoncé.*

Longueur (syll.)	Énoncés non nasaux	P_{nasaux}/P_{total}	Énoncés nasaux	P_{nasaux}/P_{total}
1	***Oui.***	0/2	***Non.***	2/2
4	***Merci beaucoup.***	1/9	***Comment vas-tu?***	2/8
10	***J'aurais voulu avoir autre chose que ça.***	0/22	***Dis-moi où je peux trouver l'homme aux cheveux bruns.***	3/24

Une fois les enregistrements effectués, les voix ont été analysées à l'aide du logiciel *Multi-Speech 3700* (modèle 3700, version 3.2.1, KayPENTAX). Une analyse générale de l'enregistrement d'un des énoncés, celui enregistré depuis un téléphone cellulaire, était effectuée afin d'obtenir la F0mp de chacune des voix ainsi que les pointes minimale et

maximale de F0 (figure 5). Les moyennes obtenues pour chacune des voix-cibles étaient de 115.79 Hertz (Hz), 134.99 Hz et 136.4 Hz (tableau IV).

Figure 5. *Exemple de courbe de la F0mp pour l'énoncé « Je suis présentement un cours de linguistique avec mon frère Jonathan » ayant servi aux mesures de comparaison de F0mp.*



3.2.1 Voix confondantes

Pour constituer les stimuli dans le groupe des voix confondantes, on a recruté des locuteurs en considérant les mêmes critères généraux que ceux utilisés pour les locuteurs présentant les voix cibles. Aucune des voix dans les deux ensembles des voix-cibles et des voix confondantes ne devait présenter une F0mp variant de plus d'un semi-ton pour une parade vocale donnée. Par ailleurs, aucun des volontaires ne devait faire partie du même réseau social qu'un participant dont la voix a été utilisée comme cible, cela afin de s'assurer que la tâche d'identification ne soit pas biaisée par la présence de plus d'une voix familière pour un participant identificateur.

La procédure ainsi que les instruments utilisés pour enregistrer les locuteurs dans les groupes de voix confondantes étaient identiques à ce qui est détaillé en 3.2 (enregistrement depuis le téléphone cellulaire) et 67 voix ont été recueillies. Les voix des volontaires ont été analysées selon la même procédure que celle utilisée pour les voix cibles en 3.2. Sur les 67 voix recueillies, 40 ont été exclues puisque, suite à l'analyse de la F0mp, la voix variait de plus d'un semi-ton de la voix-cible. Comme on le voit au Tableau V, toutes les voix

confondantes retenues se situaient à moins d'un semi-ton d'écart de la voix-cible pour une parade donnée.

Tableau V. *Valeurs absolues des différences en Hz et en semi-tons entre les voix-cibles et les voix confondantes pour chaque parade.*

Voix-cibles	**Voix confondantes**	**Différences**	
F0mp	**F0mp**	**Hz**	**Semi-ton**
136,4	135,42	0,98	0,12
136,4	131,05	5,35	0,69
136,4	136,96	0,56	0,07
136,4	129,63	6,77	0,88
136,4	135,58	0,82	0,10
136,4	129,81	6,59	0,85
136,4	138,51	2,11	0,26
136,4	135,57	0,83	0,10
136,4	130,85	5,55	0,71
115,79	115,21	0,58	0,08
115,79	109,47	6,32	0,97
115,79	111,35	4,44	0,67
115,79	122,47	6,68	0,97
115,79	116,06	0,27	0,04
115,79	109,69	6,1	0,93
115,79	114,85	0,94	0,14
115,79	119,91	4,12	0,60
115,79	117,09	1,3	0,19
134,99	131,05	3,94	0,51
134,99	136,96	1,97	0,25
134,99	129,63	5,36	0,70
134,99	135,58	0,59	0,07
134,99	128,06	6,93	0,91
134,99	129,81	5,18	0,67
134,99	138,51	3,52	0,44
134,99	135,57	0,58	0,07
134,99	130,85	4,14	0,53
	Moyenne	**(3,43)**	**(0,47)**
	Écart-type	**(2,43)**	**(0,34)**

Notons qu'afin d'obtenir un débit et une intonation similaire dans les stimuli, les locuteurs présentant les voix confondantes retenues ont premièrement écouté les

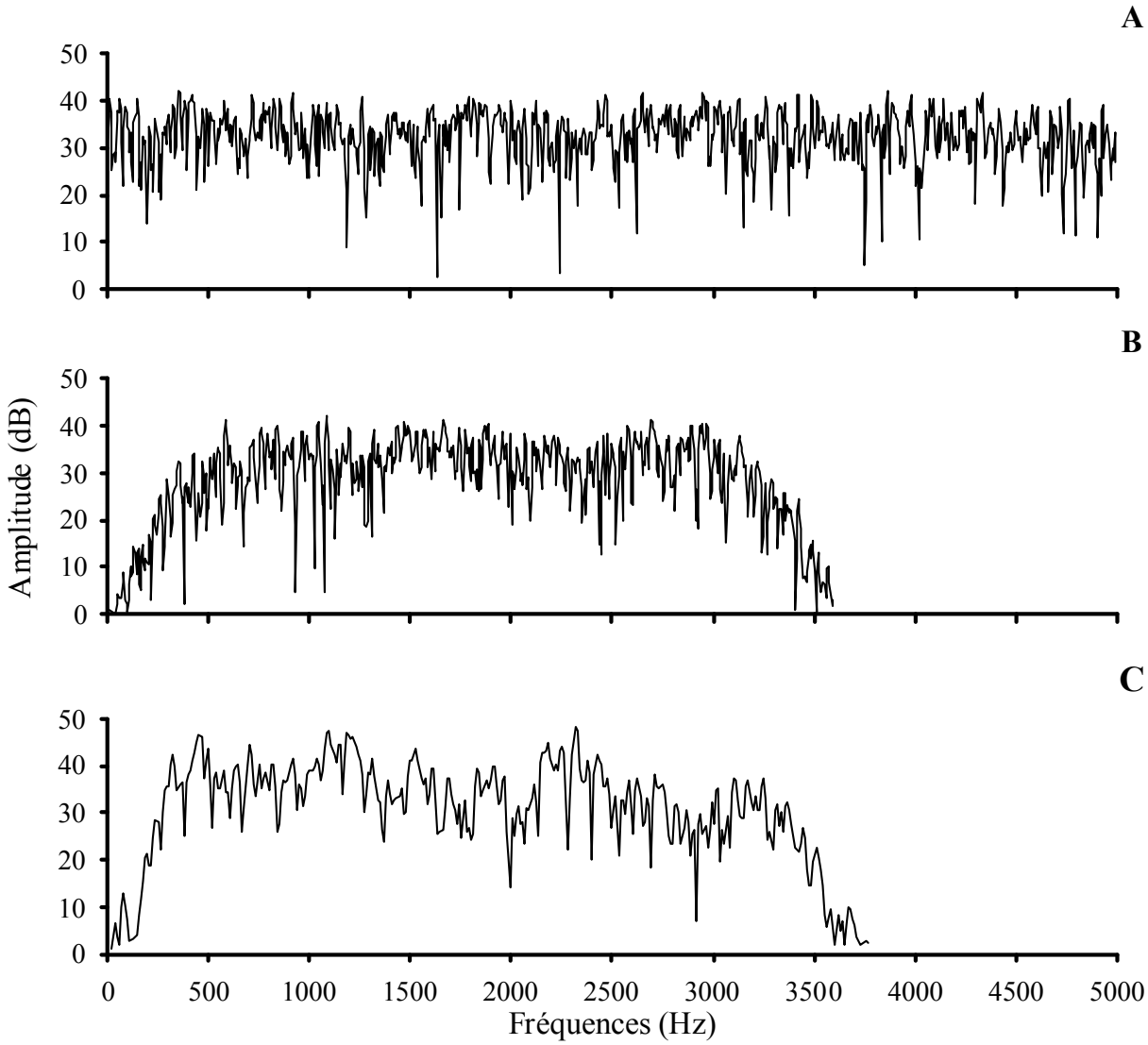
enregistrements de la voix-cible de la parade dans laquelle leur propre voix serait utilisée. Ils recevaient par la suite l'instruction de répéter, selon le débit, le rythme et l'intonation entendus, les énoncés un à un. L'expérimentateur pouvait, au besoin, demander au volontaire de répéter plus d'une fois un même énoncé si le rythme, l'intonation ou le débit différait de l'enregistrement de la voix-cible.

3.2.1.1 Modifications acoustiques des stimuli

Pour des raisons de généralisation des résultats à des situations réalistes, tous les fichiers sonores recueillis ont été modifiés. En particulier, pour reproduire les conditions de communication par cellulaire, tous les fichiers sonores enregistrés ont été filtrés de façon numérique au moyen du logiciel *Multi-Speech 3700* (modèle 3700, version 3.2.1, KayPENTAX). On a appliqué un filtre *blackman* passe-bande reproduisant les courbes de téléphone cellulaire avec des fréquences de coupe de 300 Hz à 3500 Hz.

Pour illustrer l'effet de filtrage, un bruit blanc a été envoyé dans un téléphone cellulaire (*LG 3200*) et enregistré sur un ordinateur. Une analyse comparative (figure 6) entre le bruit blanc initialement généré et l'enregistrement obtenu a été effectuée à l'aide de *Multi Speech* afin de déterminer les largeurs de bandes non retransmises par la communication cellulaire. Comme on le voit, les courbes de fréquences obtenues suite au filtre utilisé (B) sont similaires à celles obtenues lorsque le bruit blanc était directement envoyé dans le téléphone cellulaire (C). De plus, un bruit de fond audible reproduisant le bruit d'un ventilateur d'ordinateur portable et d'une largeur de bande de 0.25 kHz à 6.5 kHz avec une amplitude maximale de 24 décibels (dB) a été ajouté aux enregistrements. Celui-ci n'altérait pas la perception de la parole.

Figure 6. *Comparaison entre un bruit blanc (A,) un bruit blanc filtré (B) et un bruit blanc à travers le téléphone cellulaire utilisé (C) selon les procédures de filtrage utilisées pour constituer les stimuli audio des parades vocales. Le filtrage (B) reproduit une courbe de réponse d'un cellulaire typique (C).*



3.3 Procédure : la tâche d'identification

Pour la tâche d'identification, chacun des 44 participants était placé dans une pièce silencieuse et devait écouter au moyen d'un casque d'écoute (*Beyerdynamic*, modèle DT250) les enregistrements qui étaient joués au moyen d'un ordinateur portable (*Sony,* modèle Vaio NW Series) à une amplitude maximale moyenne d'environ 69 dBa (un niveau confortable

d'écoute) tel que mesurée avec un sonomètre et un adaptateur (*Digital Recordings*, modèle DR-1). Les directives, présentées verbalement, étaient d'écouter les 10 voix de la parade vocale présentant un même énoncé. Chaque voix était associée à un numéro lors de la présentation en format *PowerPoint* (*Microsoft*) et le participant pouvait prendre des notes sur une feuille qui lui était fournie. Après avoir entendu toutes les voix, le participant devait écrire le numéro de la voix correspondant à l'individu qu'il connaissait sur un formulaire de réponse. Suite à cette première réponse, le participant pouvait, si désiré, réécouter une ou plusieurs voix afin de confirmer ou de modifier sa réponse. Toutes les répétitions étaient notées par l'expérimentateur. Une fois la réponse validée ou modifiée, le participant procédait de la même manière pour un autre énoncé. Les énoncés étaient présentés en ordre croissant de durée (de longueur).

4 Résultats

Les résultats présentés dans cette section sont ceux obtenus après avoir offert aux participants (n = 44) la possibilité de réentendre les énoncés désirés – ce qui rejoint aussi les méthodes de la parade visuelle. En fait, lors d'une parade visuelle, un identificateur peut, au besoin, observer plusieurs fois et pour une durée relativement longue un ou plusieurs des individus présentés. Rappelons que chaque participant pouvait identifier la voix-cible 8 fois (une fois par énoncé) ce qui donnait, pour l'ensemble du groupe de 44 participants, un total de N = 352 essais.

Les coefficients présentés dans le tableau VI illustrent une forte corrélation entre le taux d'identification et le pointage obtenu pour chacun des paramètres utilisés au questionnaire portant sur la familiarité. Comme on le voit, à l'exclusion de la *récence*, les indices de *fréquence,* de *durée* et *de période* des communications avec le locuteur associé à la « voix-cible » corrèlent positivement avec l'identification exacte, ce qui, globalement, appuie à la fois la justesse de certaines mesures opérationnelles de la familiarité et l'effet de la familiarité sur l'identification des locuteurs.

Tableau VI. *Coefficients de corrélation entre les indices de familiarité et les taux d'identification. (n= 44)*

Paramètres	r_s	p
Récence	-0.052	n.s.
Fréquence	0.542	< 0.001
Durée	0.494	< 0.01
Période	0.686	< 0.001

4.1 Familiarité et longueur des stimuli

Pour mieux établir les conditions de l'effet de familiarité, on a regroupé les participants selon quatre niveaux de l'index de familiarité avec une échelle où un degré maximal de familiarité (niveau 4) était associé à des communications orales avec le locuteur de la voix-cible qui remplissait trois des quatre critères suivants :

i. *Très récentes*
ii. *Très fréquentes*
iii. *De très longues durées*
iv. *Sur une très longue période*

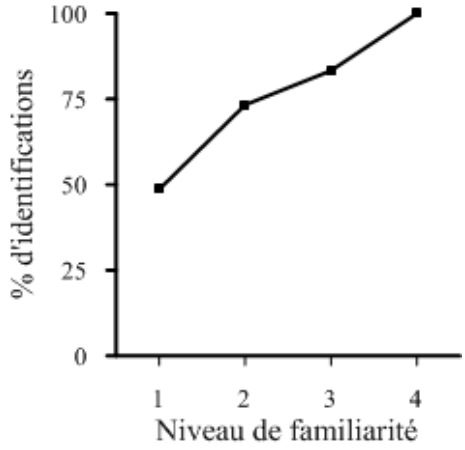
Avec cette échelle, par exemple, un niveau 3 avait des contacts assez récents (dans la dernière semaine) et assez fréquents (en moyenne 1 à 4 fois par mois), mais de très longues durées (plus de 10 minutes) et sur une très longue période (plus de 5 ans).

La fonctionnalité de l'échelle de familiarité est illustrée à la figure 7 (notons qu'en raison des résultats faibles obtenus pour les énoncés monosyllabiques, ceux-ci ont été exclus des résultats représentés dans cette figure). En considérant les énoncés de 4 syllabes et plus, on voit que les participants ayant un index de familiarité 4 (élevé, $n = 12$) ont maintenu un taux de réussite à la tâche expérimentale de 100 % pour tous ces énoncés ($N = 72$). Pour expliquer davantage ces résultats, le tableau VII illustre le nombre de participants qu'il y avait pour chacun des index de familiarité ainsi que l'étendue des résultats obtenus pour ces participants.

Tableau VII. *Nombre de participants (n) et étendue des résultats pour chaque index de familiarité.*

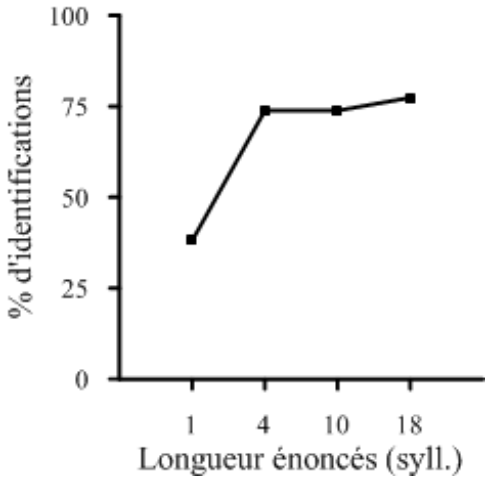
Index de familiarité	***n***	**Étendue (%)**
1	12	100
2	10	83.34
3	10	66.66
4	12	0

Figure 7. *Taux d'identification en fonction de l'index de familiarité. Pour des raisons de non-représentativité, on exclut de cette figure les résultats obtenus avec des énoncés monosyllabiques qui n'ont pas permis d'obtenir d'identifications fiables (N = 264).*



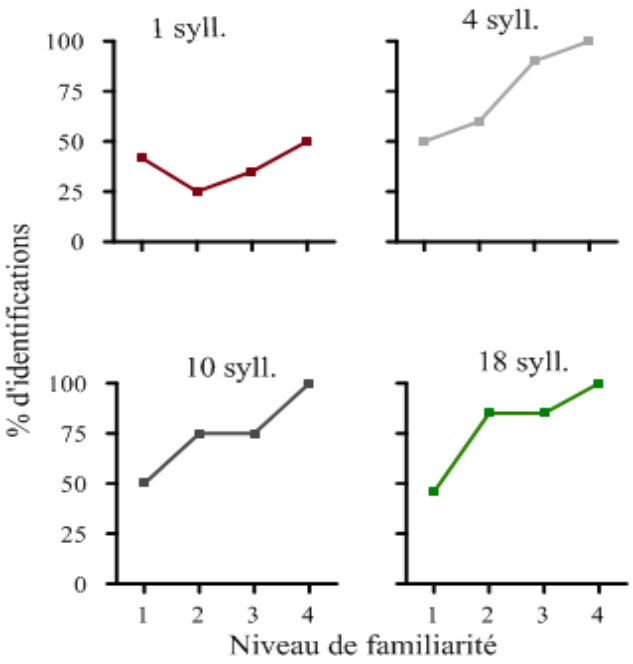
L'échelle de familiarité a permis d'isoler des effets associés aux conditions de la parade vocale dont la longueur des énoncés présentés. La figure 8 présente les résultats pour tous les participants (n = 44) obtenus en fonction de la longueur des énoncés (N = 352). On observe ici une hausse du taux d'identification en fonction de l'augmentation de la longueur des stimuli, mais celle-ci plafonne à 75 % dès une longueur de quatre syllabes.

Figure 8. *Taux d'identification en fonction de la longueur des énoncés seulement. On voit le plafonnement des taux d'identification à partir d'énoncés de 4 syll.*



Afin de mieux concevoir l'effet de longueur des énoncés, la figure 9 illustre les taux d'identification en fonction de l'index de familiarité en considérant chacune des longueurs de stimuli (N = 352). On voit qu'avec tous les énoncés de quatre syllabes et plus, on obtient des taux d'identification de 100 % pour le groupe de participants présentant un niveau de familiarité 4 avec la voix-cible.

Figure 9. *Taux d'identification en fonction de l'index de familiarité pour chaque longueur d'énoncé. Notons qu'avec un index 4 de familiarité, des identifications de locuteurs à 100 % sont obtenues avec des énoncés de 4 syll. ou plus.*



Finalement, les résultats stables obtenus par les participants avec un index de familiarité 1 à travers les différentes longueurs d'énoncés démontrent qu'il n'y a eu aucun effet d'habituation lors de la tâche expérimentale. Rappelons que ces participants n'avaient été exposés à la voix-cible qu'une seule fois pour une très courte durée peu avant l'expérimentation. La proportion de l'exposition à la voix-cible pendant la tâche expérimentale par rapport à l'exposition à la voix-cible avant la tâche est donc beaucoup plus élevée pour ce groupe de participants que pour les autres. Ils auraient donc été plus sensibles que les autres participants à un effet d'habituation lors de l'écoute des stimuli pendant la tâche expérimentale. Afin de démontrer qu'aucun tel effet n'a été observé, un test Q de Cochran a

été effectué pour les participants avec un index 1 de familiarité et indique que les différences de résultats entre les différentes longueurs d'énoncés pour ce groupe sont non significatives (Q (12) = p > 0.8).

4.2 Probabilités d'identification à 100 % pour le niveau 4 de familiarité[3]

Cette section traite des probabilités statistiques en lien avec les résultats observés pour les participants avec un index de familiarité 4 en fonction du *comportement des sommes binomiales des variables de Bernouilli* (Yu et Zelterman, 2002). Notons que toute variable dont l'issue ne peut se mesurer que par 1 (réussite) ou 0 (échec), tel que dans la présente étude, est une *variable de Bernouilli*. Dans cette étude, un total de 12 participants, toutes parades vocales confondues, avaient un index de familiarité 4. Pour ces sujets, on peut se demander quelle est la probabilité que le 100 % de reconnaissance de la voix-cible ait été obtenu par hasard.

À des fins d'illustrations, nous allons désigner ces 12 sujets par le dénoté G_4. Chaque participant, désigné ici par un indice j, devait identifier une voix dans un ensemble de 10 voix acoustiquement similaires. On tente donc de déterminer la probabilité exacte, P, que l'identification soit correcte pour un membre j du groupe G_4. Pour cela, notons d'abord une identification correcte par la valeur $X = 1$ et une fausse identification par la valeur $X = 0$, où X désigne la variable qui code l'identification (correcte ou non). Les résultats observés considèrent les énoncés des stimuli d'une longueur de 4 syllabes et plus, c'est-à-dire 6 essais par participant, ce qui se traduit par l'équation suivante pour laquelle la variable X_{kj} représente le résultat de l'identification (correct ou non, X_{kj}=0 ou 1) du k-ième essai du j-ième participant. La variable Z_j est définie comme la somme des résultats des 6 essais du j-ième participant. De façon formelle, on peut écrire :

$$Z_j = X_{1j} + X_{2j} + X_{3j} + X_{4j} + X_{5j} + X_{6j}$$

En d'autres termes, cette équation vise à déterminer combien d'essais ont été réussis pour un participant donné. Considérant que les conditions de l'expérience sont les mêmes pour tous les participants, dans la mesure où toutes les voix testées sont les mêmes et que la

[3] Cette section a bénéficié des conseils et analyses de Pr. Karim Oualkacha du département de mathématiques de l'UQAM.

longueur des énoncés n'est pas la variable responsable du taux de réussite à la tâche expérimentale, la probabilité *P* peut se calculer comme suit :

P = Probabilé (Z_j= 6 |étant donné que le j-ième sujet est dans G_4).

$$P = p_{1j} \times p_{2j} \times p_{3j} \times p_{4j} \times p_{5j} \times p_{6j}$$

Ou, pour le *k*-ième essai du *j*-ième participant, nous avons :

p_{kj} = Probabilé (X_k= 1 |étant donné que le *j*-ième sujet est dans G_4).

En vertu du design de l'expérience et considérant toujours que les conditions sont les mêmes pour chaque essai, nous pouvons supposer que :

$$p_{1j} = p_{2j} = p_{3j} = p_{4j} = p_{5j} = p_{6j} = p_j$$

C'est-à-dire que les probabilités sont identiques pour chaque essai. Ainsi, nous pouvons écrire :

$$P = {p_j}^6$$

Cependant, puisque chaque participant possède un indice de familiarité 4, ceux-ci partagent une tendance positive à performer. Ainsi, les 6 essais de chaque participant sont corrélés entre eux. Cette tendance peut être modélisée à l'aide d'un paramètre de surdispersion Θ qui, en raison de la tendance positive, sera supérieur (ou minimalement égal) à la valeur 1. Suivant Yu & Zelterman (2002), nous pouvons corriger la probabilité *P* comme suit :

$$P = {p_j}^6 \Theta^6$$

Considérant un taux d'erreur très faible (dans le cas présent, 0 % que nous approximerons à 0.01 % à des fins de calcul statistique), les probabilités que, de manière aléatoire, les 12 participants ayant un index 4 identifient la bonne voix à leurs 6 essais s'exprime de la manière suivante :

$$P = 0.01^6 1^6$$

Ou

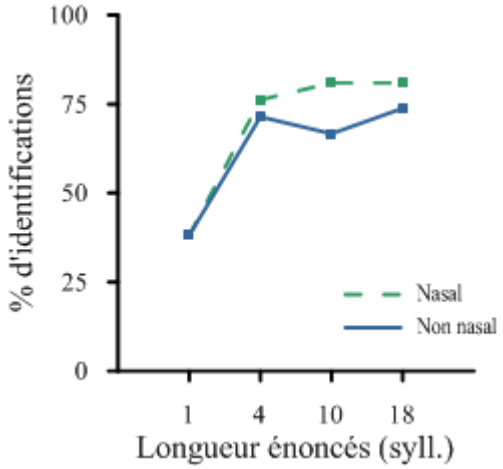
$$P = 10^{-12}$$

En somme, il est possible d'affirmer que, pour les 12 participants, la probabilité d'obtenir un taux d'identification de 100 % par pur hasard est infime, soit de 1/100,000,000,000. Notons que selon Champod et Evett (2000), une preuve est forte lorsque les probabilités d'erreur sont inférieures à 1/1000.

4.3 Effet de la nasalité

Tel que mentionné dans la section méthodologie, chaque longueur de stimuli était constituée de deux types d'énoncés variant en terme du nombre d'unités transcrites qui étaient nasales. La figure 10 présente la différence entre les taux de réussite à la tâche expérimentale pour ces deux stimuli selon chaque longueur d'énoncé ($N = 352$).

Figure 10. *Taux d'identification selon la longueur des stimuli pour les énoncés nasaux et non nasaux. Notons que la présence de sons nasaux peut contribuer à l'identification des voix à partir d'énoncés de quatre syll. et plus*.



On remarque que les données indiquent que la voix-cible est plus facilement identifiée avec les stimuli qui impliquent une plus grande utilisation des cavités nasales et que cet effet

est présent pour des énoncés de quatre syllabes et plus. Un premier test de *t* a démontré que l'écart observé entre les taux d'identification pour l'ensemble des données était non significatif [t (44) = 1.755, $p < 0.086$]. Tel que le démontre la figure 10, aucune différence entre les taux de réussite n'a été observée pour les énoncés monosyllabiques. En excluant ce cas, cependant, un second test de *t*, a révélé une différence significative [t (44) = 2.507, $p < 0.016$].

5 Discussion

L'expérience présentée dans les sections précédentes a permis d'observer la pertinence de certaines variables-clés qu'il convient de considérer dans l'élaboration d'une parade vocale. L'objectif principal était d'évaluer l'effet de la familiarité sur l'identification de locuteurs : est-ce que le fait d'avoir eu des contacts prolongés avec un individu favorise l'identification exacte de la voix de cet individu même dans un contexte où on entend les voix par bande passante reflétant celle d'un cellulaire? De façon secondaire, on voulait aussi établir l'effet de la longueur des énoncés et de la nasalité sur l'identification des voix.

Pour évaluer ces effets, on a élaboré 3 ensembles de parades vocales. Chaque ensemble contenait différentes longueurs d'énoncés prononcés par 10 voix dont une cible. Les participants, qui variaient en degré de familiarité avec la voix à identifier pour un ensemble de parades, devaient écouter toutes les voix d'une parade et identifier la voix qui leur était familière. Comme la figure 7 l'indique, les résultats démontrent que le niveau de familiarité permet effectivement une augmentation des taux d'identification correcte et que ceux-ci atteignent 100 % de réussite (pour un index 4 de familiarité). Les probabilités que ce taux soit obtenu par hasard pour tous les 12 participants ayant un index 4 de familiarité et pour les 3 longueurs de stimuli considérées sont de l'ordre de $p < .000000000001$. Ces résultats permettent d'affirmer qu'un haut niveau de familiarité entre un identificateur et un locuteur est un facteur qui permet à un individu avec une ouïe normale d'identifier une voix « hors de tout doute raisonnable ». De surcroît, les résultats illustrent qu'il est possible d'élaborer des critères opérationnels afin de quantifier la familiarité et, sur ce plan, la présente étude se distingue par rapport à certaines classifications axées sur des notions sociales et culturelles non quantifiables (Foulkes et Barron, 2000; Hollien, 1990; Yarmey et coll., 2001). Par contre, on a observé que certains facteurs associés à la familiarité, notamment la *récence* des contacts entre individus, avait peu de pertinence en ce qui a trait à l'identification des voix, comparativement à la *fréquence*, la *durée* et la *période* des contacts (voir tableau VI). Hormis la *récence*, donc, il est possible de construire une échelle de familiarité qui permet d'utiliser les capacités humaines d'identifier des voix avec exactitude. L'élaboration d'une telle échelle à l'aide de critères opérationnels permettrait, au niveau légal, de reproduire une parade vocale dans des conditions

contrôlées en ce qui a trait aux individus chargés d'identifier un suspect. Une telle procédure pourrait inciter les tribunaux à avoir recours à l'expertise de phonéticiens experts.

En ce qui a trait aux effets des facteurs secondaires, les résultats démontrent que la longueur des énoncés peut permettre une amélioration des taux d'identification, mais que ces effets bénéfiques plafonnent rapidement au-delà de quelques syllabes. En fait, dans chaque ensemble de parades vocales, on a présenté des énoncés de longueurs variables. Les résultats présentés à la figure 8 démontrent un effet de la longueur des stimuli qui plafonne aux environs de 75 % d'identifications pour des énoncés de quatre syllabes ou plus. Autrement dit, contrairement à certaines affirmations dans la littérature (p. ex. Hollien (1990)), les résultats ci-dessus indiquent que la durée d'exposition à un stimulus n'est pas suffisante pour obtenir une identification hors de tout doute raisonnable. Par contre, les observations indiquent aussi qu'un énoncé monosyllabique est insuffisant pour obtenir une identification des voix peu importe la familiarité de celles-ci. Nos résultats, sans arriver à définir une valeur exacte quant à la longueur du stimulus verbal nécessaire à une identification du locuteur, permettent de spécifier une longueur minimale propre à une identification fiable.

Enfin, l'expérience ci-dessus incorporait aussi des stimuli qui variaient en nombre de sons nasaux. Cela a permis d'observer que l'utilisation des cavités nasales influence significativement l'identification par la voix, ce qui confirme un effet bien reconnu dans la littérature (voir p.ex. Glenn et Kleiner (1967)). Aussi, l'augmentation des taux d'identification en présence de sons nasaux permet d'appuyer les conclusions avancées par Bricker et Pruzansky (1966) et Pollack et coll. (1954) selon lesquelles la taille de l'inventaire phonémique est le facteur, lié à l'augmentation de la durée d'un stimulus, permettant une hausse des taux d'identification. Ces conclusions pourraient cependant gagner à être reformulées afin de souligner qu'un plus grand nombre de configurations articulatoires, allant jusqu'à l'utilisation de plus de cavités (les cavités nasales), est sous-jacent à une augmentation de l'inventaire phonémique.

Globalement, les résultats de l'expérience ci-dessus présentent une des premières démonstrations de l'effet de la familiarité des voix, un effet souvent mentionné dans la recherche sur la reconnaissance des locuteurs, mais jamais exploité en termes de critère

opérationnel. La raison de ce désintérêt se situe dans l'idée populaire que les techniques automatisées peuvent arriver à de hauts taux d'identification. On a souligné en introduction les limites et les exigences de techniques automatisées. Les présents résultats démontrent que la parade vocale peut permettre d'obtenir des taux d'identification supérieurs à toute technique automatisée existante - et même des taux de 100 % - en contrôlant le facteur humain de familiarité et quelques aspects secondaires pouvant favoriser la perception. On pourrait objecter que les progrès de l'informatique arriveront peut-être un jour à une identification de 100 %, mais pour cela il semblerait qu'il faille incorporer des éléments de familiarisation et d'acquisition selon une *fréquence*, une *durée* et une *période* d'exposition à des voix. Par ailleurs, la rentabilité de concevoir une technologie pouvant automatiser ces aspects de familiarisation serait à évaluer par rapport à la relative efficacité de la perception humaine tel que démontré dans la présente étude.

6 Conclusion et prospective de recherche.

Les conclusions de cette étude soulèvent certaines implications et un nombre de prospectives de recherche. Tout d'abord, l'important rôle joué par le niveau de la familiarité se doit d'être considéré avec plus de précautions lors de la mise en application de la parade vocale en contexte judiciaire. Puisque les cas ayant recours à cette technique relèvent majoritairement du droit criminel, les conséquences d'une utilisation erronée de celle-ci peuvent être sérieuses. De la même manière, les effets de la durée du stimulus, comme ceux de la nasalité des sons présents dans les énoncés utilisés, nécessitent, eux aussi, une attention particulière. Il est sans contredit que les nombreuses autres variables contrôlées dans la méthodologie de la présente étude sont fondamentales dans l'élaboration d'une parade.

L'analyse des résultats et les conclusions que cette expérience apporte présentent cependant certaines limites dans le cadre de leur application. En effet, bien qu'il soit démontré qu'un individu peut, dans les conditions facilement généralisables utilisées dans la présente étude, identifier avec précision une voix très familière, il peut être difficile, voire impossible, d'obtenir une identification confirmée par des participants qui se connaissent. En d'autres termes, bien qu'un individu puisse identifier une voix connue, il se peut que cet individu n'ait pas la volonté d'identifier la voix. L'exemple d'un parent ayant comme tâche d'incriminer un fils ou une fille en l'identifiant illustre les possibles contre-indications de techniques humaines d'identification : est-ce qu'un individu ne serait pas davantage motivé à cacher la vérité ou même jeter une certaine confusion sur une enquête lorsque la voix entendue est celle d'un proche ? Ces limitations de l'identification par parade vocale, bien que sérieuses, ne sont pas sans issues si on considère que l'identification de locuteurs n'implique pas nécessairement des réponses verbales de la part de l'identificateur. En fait, ce problème motivera nos recherches futures sur les réactions physiologiques préverbales qui accompagnent la reconnaissance d'une voix familière. Si de telles réactions peuvent être observées de manière systématique, la preuve d'une identification humaine hors de tout doute raisonnable serait possible sans égard aux réponses orales ou écrites.

7 Bibliographie

Abo-Sahhad, M., Ahmed, S. M., & Abbas, S. N. (2014). Biometric authentication based on PCG and ECG signals: present status and future directions. *Signal, Image and Video Processing, 8*, 739-751.

Adibi, S. (2014). A low overhead scaled equalized harmonic-based voice authentication system. *Telematics and Informatics, 31*, 137-152.

Al-Hudhud, G., Abdulaziz Alzamel, M., Alattas, E., & Alwabil, A. (2014). Using brain signals patterns for biometric identity verification systems. *Computers in Human Behavior, 31*, 224-229.

Amino, K., Osanai, T., Kamada, T., Makinae, H., & Arai, T. (2012a). The effects of the Phonological Contents and Transmission Channel on Forensic Speaker Recognition. Dans A. Neustein & H. A. Patil (dir.), *Forensic Speaker Recognition. Law Enforcement and Counter-Terrorism* (p. 275-308). New York: Springer.

Amino, K., Osanai, T., Kamada, T., Makinae, H., & Arai, T. (2012b). Historical and Procedural Overview of Forensic Speaker Recognitions as a Science. Dans A. Neustein & H. A. Patil (dir.), *Forensic Speaker Recognition. Law Enforcement and Counter-Terrorism* (p. 3-20). New York: Springer.

Betancourt, K. S., & Huntley Bahr, R. (2010). The Influence of signal complexity on speaker identification. *The International Journal of Speech, Language and the Law, 17*, 179-200.

Boe, J. L. (2000). *Forensic voice identification in France.* Communication présentée à Institut de la Communication Parlée, Grenoble, France.

Braun, A. (1996). Age estimation by different listener groups. *Forensic linguistics, 3*, 65-73.

Braun, A., & Künzel, H. J. (1998). Is forensic speaker identification unethical - or can it be unethical not to do it? *Forensic linguistics, 5*, 10-21.

Bricker, P. D., & Pruzansky, S. (1966). Effects of Stimulus Content and Duration on Talker Identification. *The Journal of the Acoustical Society of America, 40*, 1441-1449.

Broeders, A. P. A., & van Amelsvoort, A. G. (1999). *Lineup construction for forensic earwitness identification: A practical approach.* Communication présentée à 14th International Congress of Phonetic Sciences, San Francisco, CA.

Butcher, A. (1996). *Getting the voice lineup right: analysis of a multiple auditory confrontation.* Communication présentée à the Sixth Australian International Conference on Speech Science and Technology, Adelaide, Australia.

Champod, C., & Evett, W. I. (2000). The forensic science service. *Forensic linguistics, 7*, 238-243.

Dilley, L. C., Wieland, E. A., Gamache, J. L., Devin McAuley, J., & Redford, M. A. (2013). Age-Related Changes to Spectral Voice Characteristics Affect Judgments of Prosodic, Segmental, and Talker Attributes for Child and Adult Speech. *Journal of Speech, Language, and Hearing Research, 56*, 159-177.

Donaghue, E. (2013). George Zimmerman trial: Voice experts testifies during second week of testimony. *CBS News*. Repéré le 03 décembre 2015 à http://www.cbsnews.com/news/george-zimmerman-trial-voice-expert-testifies-during-second-week-of-testimony/

Eriksson, E. J. (2007). *That voice sounds familiar. Factors in speaker recognition.* (Umeå University, Suède).

Farrús, M. (2009). Fusing prosodic and acoustic information for speaker recognition. *The International Journal of Speech, Language and the Law, 16*, 169-171.

Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *Forensic linguistics, 7*, 180-198.

French, P. (1994). An overview of forensic phonetics with particular reference to speaker identification. *Forensic linguistics, 1*, 144-153.

Glenn, J. W., & Kleiner, N. (1967). Speaker Identification Based on Nasal Phonation. *The Journal of the Acoustical Society of America, 43*, 368-372.

Gonzalez-Rodriguez, J., Torre Toledano, D., & Ortega-Garcia, J. (2007). Voice Biometrics (*Handbook of biometrics* (p. 151-169): Springer.

Harrison, P. (2001). GSM interference cancellation for forensic audio: a report on work in progress. *Forensic linguistics, 8*, 9-23.

Hollien, H. (1990). *The acoustics of crime*. New York: Springer.

Hollien, H., Hollien, P., & de Jong, G. (1997). Effects of three parameters on speaking fundamental frequency. *Journal of Acoustical society of America, 102*, 2984-2992.

Hollien, H., Huntley Bahr, R., & Harnsberger, J. D. (2014). Issues in Forensic Voice. *Journal of Voice, 28*, 170-184.

Hollien, H., Huntley Bahr, R., Künzel, H. J., & Hollien, P. (1995). Criteria for earwitness lineups. *Forensic linguistics, 2*, 143-153.

Hollien, H., & Schwartz, R. (2000). Aural-perceptual speaker identification: problems with noncontemporary samples. *Forensic linguistics, 7*, 199-211.

Horii, Y., & Ryan, W. J. (1981). Fundamental frequency characteristics and perceived age of adult male speakers. *Folia Phoniatrica, 33*, 227-233.

Hudson, T., de Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007, Aout). *F0 Statistics for 100 young male speakers of standard southern british english.* Communication présentée à International Congress of Phonetic Sciences, Saarbrücken.

Interpol (2001, Octobre). *Forensic speech and audio analysis forensic linguistics.* Communication présentée à 13th INTERPOL Forensic Science Symposium, Lyon, France.

Jain, A. K., Flynn, P., & Ross, A. A. (2007). *Biometric authetification*. USA: Springer.

Jessen, M. (2008). Forensic Phonetics. *Language and linguistics compass, 2*, 671-711.

Kanagasundaram, A., Dean, D., Sridharan, S., McLaren, M., & Vogt, R. (2014). I-vector based speaker recognition using advanced channel compendation techniques. *Computer, Speech and Laguage, 28*, 121-140.

Kerstholt, J. H., Jansen, N. J. M., van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of Accent, Retention and Telephone. *Applied Cognitive Psychology, 20*, 187-197.

Kinoshita, Y., Ishihara, S., & Rose, P. (2009). Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *The International Journal of Speech, Language and the Law, 16*, 91-111.

Köster, O., Hess, M. M., Schiller, N. O., & Künzel, H. J. (1998). The correlation between

auditory speech sensitivity and speaker recognition ability. *Forensic linguistics, 5*, 22-32.

Künzel, H. J. (1994). On the problem of speaker identification by victims and witnesses. *Forensic Linguistics, 1*, 45-57.

Künzel, H. J. (2001). Beware of the 'telephone effect': the influence of telephone transmission on the measurement of formant frequencies. *Forensic linguistics, 8*, 80-98.

Liu, P., Chen, Z., Jones, J. A., Wang, E. Q., Chen, S., Huang, D., & Liu, H. (2013). Developmental sex-specific change in auditory–vocal integration: ERP evidence in children. *Clinical Neurophysiology, 124*, 503-513.

McGehee, F. (1937). The reliability of the Identification of the Human Voice. *The journal of General Psychology, 17*, 249-271.

McGehee, F. (1944). An Experimental Study of Voice Recognition. *The journal of General Psychology, 31*, 53-65.

McLaren, M., Lawson, A., Ferrer, L., Scheffer, N., & Lei, Y. (2014, Juin). *Trial-Based Calibration for Speaker Recognition in Unseen Conditions*. Communication présentée à Odyssey 2014: The Speaker and Language Recognition Workshop, Finland.

Meinerz, C., & Masthoff, H. (2011). *Effect of Telephone-Line Transmission and Digital Audio Format on Formant Tracking Measurements*. Communication présentée à 20th International Association of Forensic Phonetics and Acoustics, Vienne.

Neustein, A., & Patil, H. A. (2012). *Forensic Speaker Recognition Law Enforcement and Counter-Terrorism*. New York, USA.

Nieman, G. S., & Applegate, J. A. (1990). Accuracy of listener judgements of perceived age relative to chronological age in adults. *Folia Phoniatrica, 42*, 327-330.

Nolan, F. (2002). The 'telephone effect' on formants: a response. *Forensic linguistics, 9*, 74-82.

Nolan, F. (2003). A recent voice parade. *Forensic linguistics, 10*, 277-291.

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007a). Earwitness Identification Performance: The Effect of Language, Target, Deliberate Strategies and Indirect Measures. *Applied Cognitive Psychology, 21*, 539-550.

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007b). Lay People's and Police Officers' Attitudes Towards the Usefulness of Perpetrator Voice Identification. *Applied Cognitive Psychology, 21*, 103-115.

Philippon, A. C., Cherryman, J., Vrij, A., & Bull, R. (2008). Why is my Voice so Easily Recognized in Identity Parades? Influence of First Impressions on Voice Identification. *Psychiatry, Psychology and Law, 15*, 70-77.

Pollack, I., Pickett, J. M., & Sumby, W. H. (1954). On Identification of Speakers by Voice. *The Journal of the Acoustical Society of America, 26*, 403-406.

Ramig, L. A., Scherer, R. C., & Titze, I. R. (1985). The aging voice. *Symposium on Care of the Professional Voice*, 1-10.

Ryan, W. J., & Burk, K. W. (1974). Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication disorders, 7*, 181-192.

Sandmann, K., am Zehnhoff-Diennesen, A., Claus-Michael, S., Rosslau, K., Lang-Roth, R., Burgmer, M., Knief, A., Matulat, P., Vauth, M., & Deuster, D. (2014). Differences Between Self-Assessment and External Rating of Voice With Regard to Sex Characteristics, Age, and Attractiveness. *Journal of Voice, 28*.

Scharinger, M., Manahan, P. J., & Idsari, W. J. (2011). You had me at "Hello": Rapid extraction of dialect information from spoken words. *Neuroimage, 56*, 2329-2338.

Shipp, T., & Hollien, H. (1969). Perception of the aging male voice. *Journal of Speech and Hearing Research, 12*, 703-710.

Su, L.-S., Li, K.-P., & Fu, K. S. (1974). Identification of speakers by use of nasal coarticulation. *The Journal of the Acoustical Society of America, 56*, 1876-1883.

Unar, J. A., Seng, W. C., & Abbasi, A. (2014). A review of biometric technology along with trends and prospects. *Pattern Recognition, 47*, 2673-2688.

Wang, L., Chen, K., & Chi, H. (2002). Capture Interspeaker Information With a Neural Network for Speaker Identification. *IEEE Transactions on neural networks, 13*, 436-445.

Wilding, J., Cook, S., & Davis, J. (2000). Sound familiar? *Psychologist, 13,* 558-562.

Yarmey, D. A. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law, 1*, 792-816.

Yarmey, D. A. (2001). Earwitness descriptions and speaker identification. *Forensic linguistics, 8*, 113-122.

Yarmey, D. A., Yarmey, L. A., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefsand the identification of familiar voices. *Applied Cognitive Psychology, 15*, 183-299.

Yu, C., & Zelterman, D. (2002). Sums of dependent Bernoulli random variables and disease clustering. *Statistics and probability letters, 57*, 363-373.