PTSVOX : une base de données pour la comparaison de voix dans le cadre judiciaire

Anaïs Chanclu, Laurianne Georgeton, Corinne Fredouille, Jean-François Bonastre

Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

PTSVOX : une base de données pour la comparaison de voix dans le cadre judiciaire

Anaïs Chanclu[1] Laurianne Georgeton[2] Corinne Fredouille[1] Jean-François Bonastre[1]
(1) LIA - Avignon Université
(2) SCPTS - Police Nationale
{anais.chanclu, corinne.fredouille,
jean-francois.bonastre}@univ-avignon.fr,
laurianne.georgeton@interieur.gouv.fr

RÉSUMÉ

Cet article présente la base de données PTSVOX, créée par le Service Central de la Police Technique et Scientifique (SCPTS) spécifiquement pour la comparaison de voix dans le cadre judiciaire. PTSVOX contient 369 locuteurs et locutrices qui ont été enregistrés au microphone et au téléphone. PTSVOX a été conçue pour mesurer l'influence de différents facteurs de variabilité fréquemment rencontrés dans les cas pratiques en identification judiciaire, comme le type de parole, le temps écoulé et le matériel d'enregistrement. Pour cela, 24 des locuteurs de PTSVOX (12 hommes et 12 femmes) ont été enregistrés une fois par mois pendant 3 mois, en parole spontanée et en parole lue. Dans cet article, nous présentons dans un premier temps la base PTSVOX, puis nous décrivons des protocoles standards ainsi que les systèmes de référence associés à PTSVOX, avec une évaluation de leur performance.

ABSTRACT

PTSVOX : a Speech Database for Forensic Voice Comparison

This article introduces PTSVOX, a forensic voice comparison database created by the Service Central de la Police Technique et Scientifique (SCPTS). PTSVOX consists in 369 speakers, recorded using a microphone, and a telephone. The database has been conceived with the purpose of studying the influence of various variability factors which are commonly encountered in practical cases, such as the speaking style, the elapsed time between two recordings and the recording equipment. This is why 24 speakers(12 women and 12 men) have been recorded once a month for three months skeaking spontaneously and reading. In this article, we first present the PTSVOX database. Then, we describe the standard protocoles and the baselines associated with PTSVOX as well as an evaluation of the performance.

MOTS-CLÉS : comparaison de voix dans le cadre judiciaire, variabilité intra-locuteur, reconnaissance du locuteur, base de données PTSVOX.

KEYWORDS: forensic voice comparison, intra-speaker variability, speaker recognition, PTSVOX database.

Volume 1 : Journées d'Études sur la Parole, pages 73–81.
HAL : hal-02798519.

1 Introduction

La reconnaissance du locuteur a réalisé des progrès importants au cours des dernières décennies, notamment grâce à la mise à disposition de grandes bases de données à travers des campagnes d'évaluations comme les campagnes NIST-*Speaker Recognition Evaluation* (SRE) organisées par le *National Institute of Standards and Technology* (NIST). La variabilité inter-locuteur, à la source de la discrimination entre les locuteurs, est bien représentée dans les bases de données type NIST-SRE. Elle est cependant mélangée avec d'autres facteurs de variabilité, comme les accents régionaux, la langue parlée, le microphone ou le bruit ambiant. De nos jours, des taux d'égale erreur (*equal error rate* (EER)) en vérification du locuteur d'environ 1 % sont couramment enregistrés dans les grandes campagnes d'évaluation. Néanmoins, ces campagnes montrent des limites (Kahn *et al.*, 2010) car le nombre d'échantillons par locuteur reste faible alors que le contrôle des facteurs de variabilité est peu réalisé. La variabilité intra-locuteur est peu représentée alors qu'elle est à la fois un problème difficile à résoudre et une limite intrinsèque de la reconnaissance du locuteur. Elle concerne les caractéristiques vocales spécifiques de la personne qui parle mais également des informations telles que la langue, le style de parole, l'état émotionnel, le contenu phonétique (Ajili *et al.*, 2016a) et l'âge (Ajili *et al.*, 2016c; Kahn *et al.*, 2010). Si la mesure de performance autorisée par de telles campagnes et bases de données peut convenir pour des applications commerciales de la reconnaissance du locuteur — pour lesquelles certains facteurs de variabilité peuvent être contrôlés ou compensés — dans certaines applications, comme la comparaison de voix dans le cadre judiciaire, cela n'est pas le cas. Il est donc nécessaire de construire les bases de données et les protocoles adaptés à ce type d'applications.

La comparaison de voix dans le cadre judiciaire est une application spécifique de la reconnaissance du locuteur où deux types d'échantillons de voix sont analysés :

- la pièce de question ou trace, relative à l'enquête, qui représente l'objet de la comparaison de voix ;
- la pièce de comparaison, dont l'origine peut être multiple (mise sous écoute, audition...), peut être un enregistrement prélevé sur un suspect lors d'un entretien.

Dans le cadre judiciaire, la variabilité est une question encore plus prégnante qu'en reconnaissance du locuteur car les conditions d'enregistrements ne sont pas contrôlées, en tout cas pour la pièce de question. Alors que les ressources en vue d'étudier la variabilité inter-locuteurs ne manquent pas, très peu de ressources ont été conçues pour étudier la variabilité intra-locuteur (Ramos *et al.*, 2008; Vloed *et al.*, 2014). La base de données FABIOLE (Ajili *et al.*, 2016b) a été mise en place pour pallier ce manque, mais elle est uniquement composée d'hommes et d'enregistrements issus d'émissions de radio et de télévision, ce qui s'éloigne du contexte de la comparaison de voix judiciaire.

Cet article présente la base de données PTSVOX, qui a été créée dans le cadre du projet ANR-17-CE39-0016 VoxCrim pour correspondre spécifiquement au contexte de la comparaison de voix judiciaire. Les protocoles standards et les systèmes de référence proposés avec PTSVOX sont aussi présentés, ainsi que les résultats expérimentaux correspondants.

2 Description de la base

PTSVOX résulte de campagnes de prélèvement de voix organisées par le Service Central de la Police Technique et Scientifique (SCPTS) dans deux écoles de police situées à Chassieu et à Nîmes. 369 personnes (144 femmes et 225 hommes) ont été enregistrées. Le corpus est composé d'enregistre-

ments téléphoniques et microphoniques réalisés sous forme d'entretien. Les personnes chargées des entretiens ont reçu pour consigne de faire parler les locuteurs autant que possible avec pour objectif de favoriser la spontanéité. Le contenu phonétique et la durée des enregistrements en mode entretien sont donc variables. La majorité des locuteurs n'a été enregistrée qu'une fois alors qu'une sous-partie du corpus, composée de 12 hommes et 12 femmes, a été enregistrée à plusieurs reprises, en Octobre 2016 puis en Mars, Avril et Mai 2017. Pour cette sous partie du corpus, des enregistrements de textes lus ont été également réalisés à chaque session, à l'exception de la première.

2.1 Fichiers audio

Chaque session d'enregistrement contient au moins deux enregistrements de parole spontanée, l'un effectué au microphone et le second au téléphone. La base compte un total de 952 fichiers audio, ce qui correspond à plus de 80 heures de données.

Microphone et téléphones Un enregistreur de type H4n est utilisé, paramétré sur une fréquence d'échantillonnage à 44100 Hz, en stéréo (car il possède deux microphones), avec une résolution de 16 bits. Trois téléphones ont été utilisés pour le prélèvement de voix, un Huawei Ascend Y550 et deux Wiko Cink Slim. Nous avons utilisé l'application Call Recorder, développée par Appliqato, sous la version 4.1.1 d'Android pour enregistrer directement les fichiers audio sur l'appareil. Les enregistrements sont paramétrés sur une fréquence d'échantillonnage de 44100 Hz, en mono, avec une résolution de 16 bits. Cependant, 27 fichiers ont été enregistrés avec une fréquence d'échantillonnage de 8000 Hz.

Ré-échantillonnage des fichiers audio Pour unifier les fichiers, les enregistrements sont également proposés après un ré-échantillonage, en mono avec une fréquence d'échantillonnage à 8000 Hz, 16000 Hz ou 44100 Hz.

2.2 Transcriptions

Les enregistrements ont été transcrits manuellement en utilisant le logiciel Praat (Boersma & Weenink, 2001) pour préparer l'alignement phonétique qui consiste à segmenter le signal de parole en unités minimales, les phonèmes. Cette segmentation en phonèmes est fournie par un outil automatique d'alignement contraint par le texte, développé par le Laboratoire Informatique d'Avignon (LIA). Cet outil prend en entrée le signal de parole, accompagné d'une transcription orthographique du contenu linguistique et un lexique de mots phonétisés (pouvant comporter différentes variantes phonologiques pour un même mot) et fournit en sortie une liste de frontières (début et fin) pour chaque phonème présent dans la transcription. L'alignement phonétique a ensuite été corrigé manuellement par des réservistes citoyennes de la Police nationale. Au total, la base contient 706 560 tokens ("mots" décomposés en chaînes de caractères) et 2 104 237 phonèmes.

2.3 Locuteurs et locutrices

Tous les locuteurs, femmes et hommes, sont des étudiants de l'école de police qui ont signé un formulaire de participation et un formulaire de consentement.

Âge Une très grande majorité des locuteurs, 253, est âgée de 18 à 24. Seuls 5 locuteurs ont plus de 30 ans. Les 111 locuteurs restants ont entre 24 et 30 ans.

Langue maternelle Le français est la langue maternelle de 346 locuteurs. D'autres langues telles que le shimaoré, les créoles guyanais, guadeloupéen et réunionnais sont également mentionnées. Seize locuteurs ont également déclaré avoir le turc, le portugais, le berbère, le malgache, l'arabe, le bushi tongo, le kurde, le guinéen ou l'italien pour langue maternelle.

État de santé Avant chaque session d'enregistrement, les locuteurs devaient indiquer si leur voix était susceptible d'être affectée par une quelconque condition ou état de santé. L'interrogatoire a montré que :

— 130 locuteurs (80 hommes et 50 femmes) ont déclaré fumer ;
— 89 locuteurs ont dit être malade le jour de l'enregistrement (nez bouché, toux, mal de gorge) ;
— 20 locuteurs ont indiqué avoir eu recours à de l'orthophonie ;
— 7 locuteurs ont subi une opération dans la zone ORL.

2.4 Jeux de données

Nous découpons la base PTSVOX en fonction des locuteurs, en trois jeux de données décrits dans le tableau 1. Il n'y a aucun recoupement de locuteurs entre les trois jeux ainsi définis. Chaque enregistrement est découpé en « tours de parole » qui sont ensuite concaténés pour obtenir des $chunks$ d'une durée minimale de 30 secondes.

	Femmes	Hommes	Sessions par locuteur	$chunks$ 30 sec	
				Microphone	Téléphone
$PTSVOX_1$	100	180	1	494	472
$PTSVOX_2$	32	33	1	151	146
$PTSVOX_3$	12	12	2 à 4	194	184

TABLE 1 – Jeux de données de la base PTSVOX

3 Protocoles et systèmes

Cette section présente les protocoles standards définis pour la base PTSVOX ainsi que les deux systèmes de référence associés.

3.1 Protocoles

Dans les protocoles présentés ci-après, nous utilisons les jeux de données décrits dans la section 2.4. Nous utilisons les enregistrements originaux rééchantillonnés à 16000 Hz.

Les protocoles sont définis sous la forme de deux listes de paires de $chunks$ (extraits d'enregistrements audio) de 30 secondes, (a, b), pour la comparaison de voix. Pour les paires $target$, les $chunks$ a et b

proviennent du même locuteur quand pour les paires $nontarget$, les $chunks$ a et b ont été prononcés par des locuteurs différents mais du même sexe.

Pour créer les paires de test $target$, le jeu de données $PTSVOX_3$ est utilisé de deux façons distinctes :

1. $PTSVOX_{3a}$: tous les tests $target$ intrasession sont exclus ;
2. $PTSVOX_{3b}$: composé uniquement des tests intrasession.

Nous distinguons trois protocoles, P_2, P_{3a} et P_{3b} dont les spécificités sont présentés dans le tableau 2 :

1. P_2 : utilise le jeu de données $PTSVOX_2$;
2. P_{3a} : utilise le jeu de données $PTSVOX_{3a}$;
3. P_{3b} : utilise le jeu de données $PTSVOX_{3b}$.

	Microphone			Téléphone		
	$target$ intrasession	$target$ transsession	$nontarget$	$target$ intrasession	$target$ transsession	$nontarget$
P_2	1206	0	1206	1268	0	1268
P_{3a}	0	6484	6484	0	5338	5338
P_{3b}	2738	0	2738	2284	0	2284

TABLE 2 – Nombre de tests par protocole

Les tests $nontarget$ ont été échantillonnés aléatoirement afin de les équilibrer en nombre avec les tests $target$.

3.2 Systèmes de référence

Nous mettons en place deux systèmes de référence, que nous évaluerons avec les trois protocoles mis en place précédemment :

1. $S_{UBM-GMM}$: approche UBM-GMM ;
2. $S_{ivector}$: approche *i-vector*.

En utilisant les approches UBM-GMM et *i-vector* (Dehak *et al.*, 2010), nous construisons deux systèmes : $S_{UBM-GMM}$ et $S_{ivector}$. Ces deux systèmes sont conçus grâce au module LIA/SpkDet, partie intégrante du toolkit open-source ALIZE (Larcher *et al.*, 2013). Les paramètres acoustiques sont composés de 19 MFCC, ses dérivées et dérivées secondes. Une normalisation des paramètres est ensuite appliquée au niveau du fichier.

L'*Universal Background Model* (UBM) possède 512 composants, et est entraîné par un algorithme *Expectation Maximisation* (EM).

Dans le cas du système $S_{UBM-GMM}$, ce modèle générique est entraîné sur le jeu de données $PTSVOX_1$. Plusieurs UBM sont créés en fonction du genre et du matériel d'enregistrement.

Dans le cas du système $S_{ivector}$, ce modèle générique est entraîné sur les données des corpus ESTER (Galliano *et al.*, 2009), ETAPE (Larcher *et al.*, 2013) et REPERE (Giraudel *et al.*, 2012).

Les matrices T de variabilité totale sont également apprises sur ces données. Deux UBM et deux matrices T sont créés en fonction du genre.

Le score délivré lors d'une comparaison de voix est le logarithme du rapport de vraisemblance (*log-likelihood ratio* (LLR)) exprimé par :

$$score = log \frac{P(e_1, e_2|H_p)}{P(e_1, e_2|H_d)} \quad (1)$$

où H_p est l'hypothèse e_1 et e_2 proviennent de la même personne, et H_d est l'hypothèse que e_1 et e_2 ont été prononcés par des personnes différentes.

3.3 Performances

Les tableaux 3 et 4 indiquent les performances pour les deux systèmes de référence en fonction du protocole de test, avec une paramétrisation en bande large 0-8000 Hz. La performance est exprimée en taux d'égale erreur (EER, le taux d'erreur totale est ici le double de l'EER). Pour le système $S_{UBM-GMM}$, les résultats sont donnés pour les UBM appris sur des données téléphoniques et les UBM appris sur des données microphoniques.

		Femmes		Hommes	
		UBM Mic.	UBM Tél.	UBM Mic.	UBM Tél.
P_2	Microphone	1.32%	13.21%	0.66%	8.05%
	Téléphone	7.62%	1.07%	3.32%	1.84%
P_{3a}	Microphone	2.50%	39.02%	2.38%	34.16%
	Téléphone	14.47%	40.33%	11.80%	38.53%
P_{3b}	Microphone	0.60%	9.02%	0.31%	7.43%
	Téléphone	8.08%	6.26%	14.06%	5.18%

TABLE 3 – Performances (EER) pour le système $S_{UBM-GMM}$ avec une bande passante large 0-8000 Hz

		Femmes	Hommes
P_2	Microphone	0.00%	0.00%
	Téléphone	0.00%	0.14%
P_{3a}	Microphone	10.24%	6.44%
	Téléphone	43.52%	40.49%
P_{3b}	Microphone	4.55%	5.15%
	Téléphone	11.41%	9.27%

TABLE 4 – Performances (EER) pour le système $S_{ivector}$ avec une bande passante large 0-8000 Hz (UBM et matrice de variabilité totale T, appris sur ESTER-ETAPE-REPERE)

Les tableaux 5 et 6 indiquent les performances pour une situation comparable aux tableaux 3 et 4 mais avec une paramétrisation en bande passante étroite, 300-3400 Hz.

		Femmes		Hommes	
		UBM Mic.	UBM Tél.	UBM Mic.	UBM Tél.
P_2	Microphone	1.16%	1.99%	1.50%	2.82%
	Téléphone	4.46%	1.07%	4.38%	2.12%
P_{3a}	Microphone	7.82%	12.94%	2.18%	7.35%
	Téléphone	23.55%	20.41%	20.92%	21.47%
P_{3b}	Microphone	2.66%	6.79%	0.44%	3.37%
	Téléphone	5.56%	1.92%	6.96%	5.26%

TABLE 5 – Performances (EER) pour le système $S_{UBM-GMM}$ en bande étroite 300-3400 Hz

		Femmes	Hommes
P_2	Microphone	0.00%	0.17%
	Téléphone	0.00%	0.00%
P_{3a}	Microphone	27.16%	9.08%
	Téléphone	29.59%	32.20%
P_{3b}	Microphone	10.82%	5.33%
	Téléphone	8.28%	6.18%

TABLE 6 – Performances (EER) pour le système $S_{ivector}$ en bande étroite 300-3400 Hz (UBM et matrice de variabilité totale T, appris sur ESTER-ETAPE-REPERE)

Le tableau 7 montre la performance du système $S_{UBM-GMM}$ en utilisant l'UBM du système $S_{ivector}$, appris sur ESTER-ETAPE-REPERE, en bande étroite.

		Femmes	Hommes
P_2	Microphone	2.48%	2.16%
	Téléphone	2.16%	5.51%
P_{3a}	Microphone	17.94%	5.73%
	Téléphone	26.74%	21.93%

TABLE 7 – Performances (EER) pour le système $S_{UBM-GMM}$ en utilisant un UBM appris sur ESTER-ETAPE-REPERE, avec une bande passante étroite 300-3400 Hz

4 Discussion

Pour le protocole P_2 pour lequel il n'y a pas de paires $target$ inter-session, la performance obtenue par les deux systèmes de référence est bonne pour toutes les situations. Le système $S_{ivector}$ montre une supériorité, malgré le fait que ses données d'apprentissage ne viennent pas de PTSVOX. Pour

le protocole P_{3a}, qui ne contient que des paires inter-sessions, deux phénomènes sont observés. D'une part, la performance diminue légèrement comparativement à P_2 en données microphone pour $S_{UBM-GMM}$ et plus fortement pour $S_{ivector}$. D'autre part, pour les données téléphone, la performance s'écroule complètement pour les deux systèmes et se rapproche du hasard. Une hypothèse plausible pour expliquer ces résultats est la présence d'un facteur de variabilité important entre les différentes sessions d'enregistrement "téléphone" : les systèmes reconnaissent autant la session que le locuteur. Les résultats obtenus en utilisant le protocole P_{3b} confirment cette hypothèse : le niveau de performance est alors proche de celui obtenu pour P_2.

Il est difficile d'expliquer cette importante variabilité entre sessions téléphoniques autrement que par un artefact technique. L'analyse spectrale de plusieurs fichiers téléphoniques du protocole P_{3a} montrent une réflection du spectre entre 4000 Hz et 5000 Hz , ce qui pourrait expliquer les résultats observés précédemment. Les expériences en bande étroite confirment que si cette réflection est une part du problème, d'autres facteurs existent également : en bande 300-3400 Hz, alors que la réflexion est exclue, l'écart de performance entre P_{3a} et P_{3b} se réduit mais reste marqué, quel que soit le système employé (une expérience complémentaire avec une bande passante 0-4000 Hz a montré des résultats similaires). Les particularités ou des différences de configuration du logiciel d'enregistrement Call Recorder sont une deuxième source plausible de cet effet session. Par ailleurs, sans surprise, en données microphone, une dégradation des performances est observée quand la bande passante est réduite, dégradation logiquement beaucoup plus forte pour les voix de femmes que pour les voix d'hommes.

Le système $S_{UBM-GMM}$ utilisant un UBM appris sur d'autres données que PTSVOX (tableau 7) montre une perte de performance comparativement à un UBM appris sur PTSVOX (tableau 5) pour les données microphone mais une amélioration de la performance pour les données téléphone, ce qui renforce l'idée que les données téléphone de PTSVOX contiennent un facteur de variabilité important et rarement rencontré. Ceci explique également le très bon, voire trop bon, niveau de performance observé avec le protocole P_2 : l'information sur la session d'enregistrement est tellement marquée qu'elle est peut-être plus saillante même que l'information locuteur.

5 Conclusion et perspectives

Dans cet article, nous présentons la base PTSVOX, dédié à la comparaison de voix judiciaire. Le corpus contient des enregistrements téléphoniques et microphones de parole spontanée et de lecture provenant de 369 personnes, représentant environ 80 heures au total. La base a été transcrite orthographiquement et alignée phonétiquement. Trois sous corpus et trois protocoles expérimentaux ont été définis, ainsi que deux systèmes de comparaison de voix de référence. Les expériences menées avec ces protocoles et systèmes ont démontré la bonne qualité des systèmes de référence. Elles ont permis de repérer un effet session prononcé et non audible, indicateurs d'artefacts, pour les données téléphone. Ces expériences illustrent également la sensibilité de la comparaison de voix à des facteurs de variabilité inconnus et met en lumière la nécessité d'analyser et de comprendre les raisons d'une différence de performance, dans un sens ou l'autre.

Pour faciliter l'accès de PTSVOX, un outil développé avec MongoDB est en cours de développement. Cet outil permet de créer des protocoles de test spécifiques à travers une interface Web aisée d'accès. L'outil permettra d'automatiser l'exécution des systèmes de référence sur ces protocoles. La base PTSVOX, les systèmes de référence et les outils associés seront diffusés prochainement grâce à une licence libre.

Remerciements

Ce travail de recherche a été financé par le projet ANR-17-CE39-0016 VoxCrim qui inclut le Laboratoire Informatique d'Avignon, le Laboratoire Parole et Langage, le Laboratoire Phonétique et Phonologie, le Service Central de la Police Technique et Scientifique et l'Institut de Recherche Criminelle de la Gendarmerie Nationale. Les auteurs tiennent également à remercier Nathan Griot pour la conception de l'interface ainsi que les réservistes citoyennes de la Police nationale qui ont travaillé sur la base de données.

Références

AJILI M., BONASTRE J.-F., BEN KHEDER W., ROSSATO S. & JULIETTE K. (2016a). Phonetic content impact on forensic voice comparison. In *2016 IEEE Spoken Language Technology Workshop (SLT)*, p. 210–217.

AJILI M., BONASTRE J.-F., KAHN J., ROSSATO S. & BERNARD G. (2016b). Fabiole, a speech database for forensic speaker comparison. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, p. 726–733.

AJILI M., BONASTRE J.-F., ROSSETTO S. & KAHN J. (2016c). Inter-speaker variability in forensic voice comparison : a preliminary evaluation. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 2114–2118.

BOERSMA P. & WEENINK D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, **5**(9/10), 341–345.

DEHAK N., KENNY P. J., DEHAK R., DUMOUCHEL P. & OUELLET P. (2010). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(4), 788–798.

GALLIANO S., GRAVIER G. & CHAUBARD L. (2009). The ester 2 evaluation campaign for the rich transcription of french radio broadcasts. In *INTERSPEECH '09*. Brighton, Royaume-Uni.

GIRAUDEL A., CARRÉ M., MAPELLI V., KAHN J., GALIBERT O. & QUINTARD L. (2012). The repere corpus : a multimodal corpus for person recognition. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, p. 1102–1107.

KAHN J., AUDIBERT N., ROSSATO S. & BONASTRE J.-F. (2010). Intra-speaker variability effects on speaker verification performance. In *Odyssey*, p.2̃1.

LARCHER A., BONASTRE J.-F., FAUVE B., LEE K. A., LEVY C., LI H., MASON J. & PARFAIT J.-Y. (2013). Alize 3.0-open source toolkit for state-of-the-art speaker recognition. In *INTERSPEECH '13*. Lyon, France.

RAMOS D., GONZALEZ-RODRIGUEZ J., GONZALEZ-DOMINGUEZ J. & LUCENA-MOLINA J. J. (2008). Addressing database mismatch in forensic speaker recognition with ahumada iii : a public real-casework database in spanish. In *Ninth Annual Conference of the International Speech Communication Association*.

VLOED D., BOUTEN J. & VAN LEEUWEN D. (2014). Nfi-frits : A forensic speaker recognition database and some first experiments. In *Proceedings of Odyssey Speaker and Language Recognition Workshop*, p. 6–13 : [Sl] : ISCA Speaker and Language Characterization special interest group.