

# Description automatique du taux d’expression des femmes dans les flux télévisuels français

David Doukhan Jean Carrive

Institut national de l’audiovisuel (Ina), 4 avenue de l’Europe, 94366 Bry-sur-Marne cedex, France  
ddoukhan@ina.fr, jcarrive@ina.fr.fr

## RÉSUMÉ

---

Une approche automatique, fondée sur l’estimation du temps de parole par genre, est proposée pour décrire le taux d’expression féminine dans les flux audiovisuels. Des réseaux de neurones convolutionnels (CNN) sont utilisés pour segmenter les flux audio en zones de musique et de parole, attribuées à des hommes ou des femmes. Le taux d’expression féminine est décrit sur plus de 170 000 heures de flux, correspondant à 21 chaînes de télévision nationale, analysées de 2010 à 2017, ainsi que sur 24 chaînes de télévision régionales, analysées en 2016. Cette description montre que le temps de parole est majoritairement attribué à des hommes dans la télévision française, mais que le pourcentage de parole attribué aux femmes a significativement progressé en huit ans pour au moins sept chaînes.

## ABSTRACT

---

### Automatic description of female speaking time percentage in French TV streams

This paper presents an automatic approach, based on speaker gender detection, aimed at describing female speaking time percentage in audiovisual streams. Convolutional Neural Network models (CNN) are used to segment audio streams into music and speech segments, attributed to male and female speakers. Female speaking time percentage is described using 170.000 hours of raw streams corresponding to 21 national TV channels, analysed between 2010 and 2017, and 24 regional TV channels, analyzed in 2016. This description shows that speech time is mainly attributed to male speakers in French TV, and that female speaking time percentage has statistically increased in 8 years for 7 channels.

---

**MOTS-CLÉS :** Détection du genre du locuteur, Humanités Numériques, Égalité des sexes, ConvNet.

**KEYWORDS:** Speaker Gender Detection, Digital Humanities, Gender Equality, ConvNet.

---

## 1 Introduction

La mesure de la place accordée aux femmes dans les médias a été traitée à l’aide d’un grand nombre de méthodologies. A l’échelle européenne, elle a été décrite à l’aide du pourcentage de femmes accédant aux plus hauts postes décisionnels dans les médias par type de métier ainsi que du pourcentage de sujets féminins traités dans les nouvelles (Europe, 2015). Elle a également fait l’objet d’études réalisées à l’échelle mondiale (Macharia, 2015), de rapports publics (Reiser & Gresy, 2008), ou encore d’analyses réalisées par le Conseil supérieur de l’audiovisuel (CSA), en charge de veiller à une plus juste représentation des femmes et des hommes dans les programmes audiovisuels depuis 2014 (CSA, 2017). La place des femmes dans les médias y a été décrite à l’aide du *taux de présence* (pourcentage de femmes présentes à l’antenne), du *taux d’expression* (pourcentage de temps de parole

attribué aux femmes), ainsi que du *taux d'identification* (pourcentage de mentions orales du nom de personnalités féminines). Ces taux peuvent être exprimés de manière globale, ou analysés séparément en fonction des parts d'audience ou du statut des locuteurs : invité politique, journaliste...

La mesure manuelle du *taux d'expression* des femmes dans les médias est coûteuse et les études s'y consacrant ont jusqu'ici été réalisées sur des corpus de taille limitée. Le corpus GMM, constitué à partir d'émissions diffusées le 15 mai 2008 sur 6 chaînes de télévision et 6 stations de radio, est composé d'extraits d'une durée variant entre 6 minutes et 3 heures par chaîne (Reiser & Gresy, 2008). L'étude menée par le CSA belge rapporte des temps de parole observés sur un total de 36 heures d'archives collectées sur une période d'une semaine sur 26 chaînes de télévision (Levant *et al.*, 2014). Ces limitations provoquent des biais d'analyse correspondant aux particularités du contexte politique et social dans lequel les programmes ont été collectés. Ces études, basées sur l'analyse d'une période restreinte, sont systématiquement accompagnées d'une description détaillée des événements relatifs à cette période : élections, mouvements sociaux, manifestations sportives et culturelles... Ces éléments étant nécessaire pour caractériser le biais affectant les descripteurs de la place accordée aux hommes et aux femmes dans les médias.

L'approche proposée dans cet article consiste à décrire l'évolution du taux d'expression féminine à l'aide de systèmes de segmentation automatique, permettant d'analyser une masse de données beaucoup plus importante, et ainsi de réduire les effets de biais liés aux contexte de collecte. Ils permettent également d'étendre les analyses à l'ensemble des programmes diffusés et ainsi de mettre en évidence un certain nombre de phénomènes susceptibles d'orienter des analyses qualitatives.

La détection du genre d'un locuteur est rendue possible à l'aide d'un certain nombre de caractéristiques acoustiques. La parole des femmes est généralement caractérisée par une fréquence fondamentale plus élevée, des formants localisés dans de plus hautes plages de fréquences, ainsi qu'une qualité de voix plus *soufflée* (breathy). Les différences prosodiques existant entre la parole des hommes et des femmes varient en fonction des pays : cela signifie que les marqueurs acoustiques du genre dans la parole ne sont pas uniquement physiologiques, mais également appris dans un contexte socio-culturel donné (Pépiot, 2015). La détection du genre est plus difficile pour les locuteurs ayant un fort accent (étranger ou régional), une voix associée à des plages de fréquence fondamentale extrêmes, ou encore s'exprimant à l'aide de motifs prosodiques non standard (voix très expressive, imitation, etc...).

Les systèmes de détection du genre du locuteur sont utilisés depuis de nombreuses années dans les moteurs de transcription automatique de la parole (Lamel & Gauvain, 1995). La détection du genre n'y est pas forcément considérée comme une fin en soi, mais plutôt comme un moyen d'améliorer la qualité de la transcription en sélectionnant le modèle acoustique le plus approprié. Plus récemment, des systèmes de détection du genre ont été utilisés pour améliorer les tâches de reconnaissance d'émotions (Xia *et al.*, 2014), ainsi que les stratégies d'interaction d'interface homme-machine (El Shafey *et al.*, 2014). La détection du genre du locuteur présente encore un certain nombre de difficultés, notamment lorsqu'il s'agit de personnes âgées ou d'enfants (Schuller *et al.*, 2013).

## 2 Système de Segmentation audio

Le système proposé pour la segmentation des flux audio est composé de 3 modules principaux : un détecteur d'activité (seuil énergétique), un système de segmentation parole/musique et un système de segmentation homme/femme. Il est disponible sur le répertoire GitHub de l'Ina sous license open-

source<sup>1</sup>. Les systèmes de segmentation sont modélisés à l'aide de réseaux de neurones convolutionnels composés de 5 couches convolutionnelles et 4 couches denses. Ils utilisent le Mel-Frequency Cepstrum (MFC) du signal audio extrait avec SIDEKIT (Larcher *et al.*, 2016), en utilisant un banc de 24 filtres espacés entre 100 et 8000 Hz, ainsi qu'une taille de pas et de fenêtre de 10 et 25 millisecondes. L'entrée des réseaux de neurones est composée de fenêtres de 68 MFC (contexte de 680 millisecondes, 68\*24 dimensions) normalisées localement en soustrayant leur moyenne et en les divisant par leur écart type. La sortie des systèmes correspond à une probabilité d'émission par classe (homme-femme ou parole-musique) qui est estimée toutes les 20 millisecondes, et est fournie en entrée d'un modèle de Markov caché à deux états (HMM), chargé de transformer ces prédictions brutes en segments.

Les modèles de détection du genre ont été entraînés à l'aide du dictionnaire de locuteurs interne de l'Ina (Salmon & Vallet, 2014), qui est à notre connaissance la plus grande base de locuteurs annotée manuellement à partir de données audiovisuelles (TV et radio). Ce corpus contient environ 32 000 extraits de parole, attribués à 1 780 hommes (94 heures) et 494 femmes (27h). Ces extraits ont été obtenus à partir d'archives collectées entre 1957 et 2012. L'exhaustivité de cet ensemble d'apprentissage fournit un avantage technologique certain pour la conception de systèmes automatiques, et permet d'entraîner des modèles de reconnaissance plus complexes (nombre de paramètres libres, profondeur) en limitant les risques de sur-apprentissage.

La description détaillée et l'évaluation du système ont été réalisées lors de précédents travaux, où ses performances avaient été comparées à des systèmes fondés sur l'utilisation de *i-vecteurs* ainsi que de modèles de mélanges de gaussiennes (Doukhan *et al.*, 2018a). Le système CNN était associé à de meilleurs résultats bruts (F-Mesure de 96.52%), ainsi qu'à une meilleure capacité à estimer le taux d'expression dans des extraits de durée variant entre 30 et 60 minutes (erreur moyenne : 0.59%, erreur maximale : 1.8%). Il a également été montré que plus la durée de l'enregistrement considéré est longue, plus l'erreur d'estimation du taux d'expression est basse. Suite à cette étude, le système a été jugé suffisamment robuste pour pouvoir décrire de manière fiable les variations du taux d'expression.

## 3 Variations du taux d'expression dans les flux TV nationaux

### 3.1 Corpus d'étude

La table 1 présente les 21 chaînes sélectionnées pour décrire les variations du taux d'expression dans les flux télévisuels nationaux. Cette sélection est composée de 7 chaînes publiques et 14 chaînes privées. Elle a été réalisée de manière à faire apparaître les chaînes associées aux plus fortes audiences, ainsi que certaines chaînes thématiques cibles (informations, sport, histoire, musique, contenu visant un public féminin).

Les analyses ont porté sur l'ensemble des flux diffusés de 2010 à 2017. Afin de réduire le temps de calcul, les flux ont été découpés en tranches d'une heure, et un échantillon de 20% des segments a été sélectionné aléatoirement pour être analysé. La plage horaire d'analyse a été restreinte de 10h du matin à minuit, ce qui correspond aux créneaux pour lesquels les audiences de la télévision sont supérieur à 10% (CSA, 2015). La quantité d'échantillons analysés correspond au nombre de chaînes (21) multiplié par le nombre d'heures analysées par jour (14h), le nombre de jours (8 ans) ainsi que le taux d'échantillonnage (20%); soit environ 170 000 heures de contenu (20 ans de flux brut).

---

1. <http://github.com/ina-foss/inaSpeechSegmenter>

Code	Nom	Statut	Contenu
ART	Arte	Public	Chaîne franco-allemande à vocation culturelle
BFT	BFM TV	Privé	Information nationales en continu
C+	Canal+	Privé	Généraliste axée sur le cinéma et le sport
C25	Chérie 25	Privé	Généraliste visant un public féminin
E21	L'Équipe TV	Privé	Thématique à contenu sportif
ESP	Eurosport	Privé	Thématique à contenu sportif
F24	France 24	Public	Chaîne publique d'information internationales en continu diffusée en 4 langues et dans 180 pays
FR2	France 2	Public	Généraliste. Deuxième chaîne la plus regardée en France
FR3	France 3	Public	Généraliste à vocation régionale : 24 éditions régionales, 44 éditions locales, et 6 éditions en langues régionales
FR5	France 5	Public	Généraliste axé sur l'éducatif et les documentaires
FRO	France Ô	Public	Généraliste consacré à la France d'outre-mer
HIS	Histoire	Privé	Thématique consacré à l'histoire
ITL	I-Télé	Privé	Information nationale en continu
LCI	La Chaîne Info	Privé	Information nationale en continu
LCP	LCP/Public Sénat	Public	Politique (Assemblée nationale et Sénat), information
M6	M6	Privé	Généraliste. Troisième chaîne la plus regardée en France.
N12	NRJ 12	Privé	Généraliste axée sur les divertissements
TEV	Téva	Privé	Généraliste visant un public féminin et familial
TF1	TF1	Privé	Généraliste. Chaîne la plus regardée en Europe
TMC	TMC	Privé	Généraliste
W9	W9	Privé	Généraliste à vocation musicale et de divertissement

TABLE 1 – Chaînes de télévision sélectionnées pour décrire le taux d'expression par genre

## 3.2 Description Globale

La figure 1 présente une visualisation globale des 21 chaînes du corpus selon deux modalités : le pourcentage de parole (100 - pourcentage de musique), et le taux d'expression féminine (100 - pourcentage de parole masculine), observés entre 2010 et 2017.

Le pourcentage de paroles varie entre 61 et 93,7%. Il est minimal pour W9 (chaîne musicale) et maximal pour l'ensemble des chaînes d'information, ainsi que les chaînes de sport.

Le taux d'expression des femmes varie entre 7,73 et 47,44%, ce qui signifie que dans la totalité des chaînes analysées les hommes ont un temps de parole supérieur aux femmes. Il est minimal dans les chaînes ayant une programmation sportive (Eurosport, L'Équipe, et dans une moindre mesure CANAL+), et légèrement inférieur à la moyenne dans les chaînes à programmation culturelle (Histoire, Arte, France 5). Les chaînes d'information en continu privées (I-Télé, LCI, BFM-TV) ont des propriétés similaires (taux de parole compris entre 89,9 et 90,6%, pourcentage de parole féminine compris entre 33,4 et 36,4%). Quatre chaînes sont associées à un taux d'expression des femmes supérieur à 40% : les deux chaînes visant un public féminin (Téva et Chérie 25), France 24 et M6.

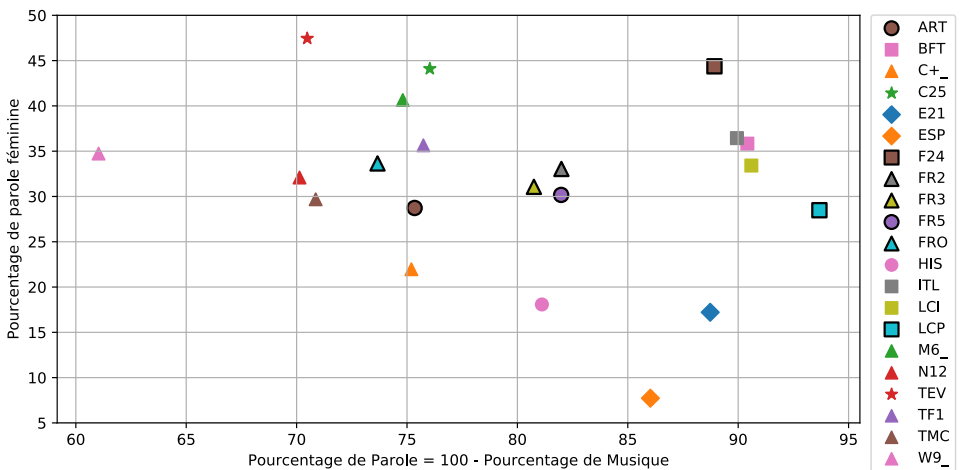


FIGURE 1 – Pourcentages de parole et taux d’expression des femmes observés à la télévision de 2010 à 2017. Les marqueurs à bordure noire correspondent aux chaînes publiques, les triangles aux chaînes généralistes, les cercles aux chaînes à contenu culturel ou éducatif, les losanges au contenu sportif, les carrés aux chaînes d’information et politiques, les étoiles aux chaînes visant un public féminin.

chaîne	FR5	W9	M6	ITL	TF1	FR2	LCP	TEV
pente	1.29	0.89	0.52	-1.16	0.67	1.05	0.68	1.29
p-score	3.9e-04	7.6e-04	2.2e-03	3.3e-03	7.3e-03	8.3e-03	8.7e-03	3.3e-02

TABLE 2 – Évolution du pourcentage de parole des femmes entre 2010 et 2017 pour les chaînes ayant une pente statistiquement significative ( $p\text{-score} < 0.05$ )

### 3.3 Évolution du taux d’expression des femmes de 2010 à 2017

La figure 2 présente l’évolution du pourcentage de parole féminine pour 9 chaînes du corpus associées à des fortes audiences. Pour l’ensemble de ces chaînes, le taux de parole féminine estimé en 2017 est supérieur à celui estimé en 2010. Les chaînes associées aux plus fortes évolutions sont France 2 qui passe de 27,7 à 37,2% et France 5 qui passe de 27,4 à 35,1%.

La table 2 présente les tendances associées cette évolution. Une procédure de régression linéaire (`scipy.stats.linregress`) a été utilisée pour associer à chaque chaîne une mesure de la pente du pourcentage de parole féminin, ainsi qu’un p-score obtenu à l’aide du test de Wald. Huit chaînes sur 21 ont été associées à une évolution significative du pourcentage de parole féminine ( $p\text{-score} < 0.05$ ). La chaîne d’information I-Télé est la seule qui est associée à une diminution du temps de parole des femmes (-1.16% par an). Les trois chaînes associées aux plus fortes croissantes du taux d’expression des femmes sont France 5, Téva, et France 2.

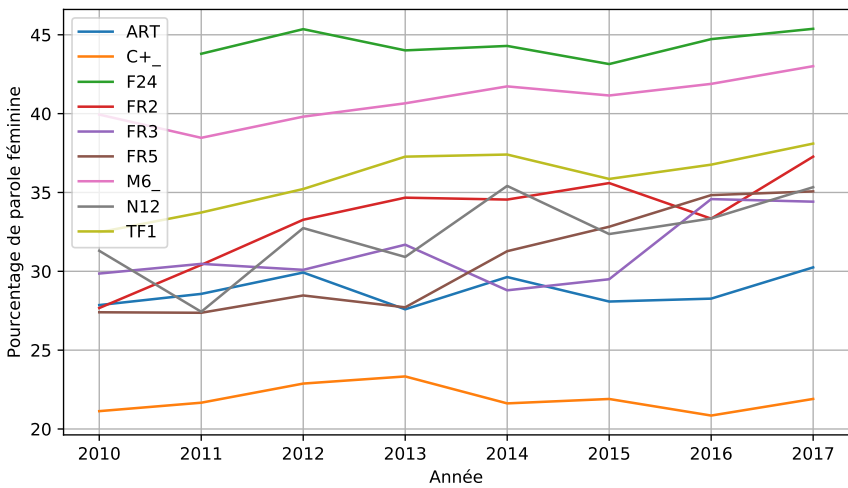


FIGURE 2 – Évolution du taux d’expression des femmes de 2010 et 2017

## 4 Variations du taux d’expression dans les flux TV régionaux

### 4.1 Corpus d’étude

L’étude des disparités régionales du taux d’expression par sexe a été réalisée en analysant la totalité des éditions régionales du 19/20 de France 3 diffusées en 2016. Le 19/20 est une émission d’information diffusée en *prime-time* sur France 3, associée à des fortes parts d’audience, pouvant varier de 14 à 21% (Ozap, 2010; Guadalupe, 2016; Meffre, 2017).

Les 24 éditions régionales du 19/20 sont diffusées simultanément sur France 3, de 19h à 19h30. Elles peuvent être entrecoupées de programmes de publicité ou de bulletins météo, et sont suivies par une édition nationale du 19/20. Les éditions régionales correspondent au découpage en 21 régions métropolitaines de la France datant d’avant 2016, auxquelles s’ajoute la Corse. La région Provence-Alpes-Côte d’Azur a deux éditions distinctes (Provence-Alpes, Côte d’Azur), ainsi que la région Rhône-Alpes (Rhône, Alpes).

### 4.2 Sélection des données

Les éditions régionales ont été isolées dans les flux à l’aide d’une méthode automatique de traitement d’image relativement simple (CNN), consistant à identifier la bannière associée à l’ensemble des journaux régionaux (figure 3). Cette stratégie permet de détecter avec précision l’heure de début et de fin de l’édition, et d’exclure de l’analyse les programmes non désirés : publicité, météo, édition spéciale diffusée sur l’ensemble des chaînes, édition nationale, émission de substitution. Chaque édition a été associée à 5 jours et demi de données brutes, soit 4 mois de flux en considérant l’ensemble des régions, qui ont été analysés dans leur intégralité.



FIGURE 3 – Capture d’écran d’une édition régionale du 19/20 sur France 3, contenant quasi-systématiquement une bannière intitulée 19/20 suivi du nom de la région ou localité considérée

### 4.3 Disparités régionales

La figure 4 détaille le pourcentage de parole attribué aux femmes dans les 24 éditions régionales du 19/20. Ce pourcentage varie entre 25.89% et 52.9%. L’Alsace et Nord-Pas-de Calais sont les seules éditions pour lesquelles le temps de parole attribué aux femmes est supérieur au temps attribué aux hommes. Sept éditions sur 24 sont associées à un temps de parole par sexe à peu près égal (compris entre 45 et 55%), à savoir Alsace (52,90%), Nord-Pas-de-Calais (50,67%), Ile-de-France (47,25%), Picardie (47,13%), Bretagne (46,34%), Provence-Alpes (46,31%), Languedoc-Roussillon (46,06%). Quatre éditions régionales sont associées à un temps de parole des femmes inférieur à un tiers : Lorraine (25,89%), Midi-Pyrénées (31,25%), Auvergne (32,72%), Aquitaine (32,91%).

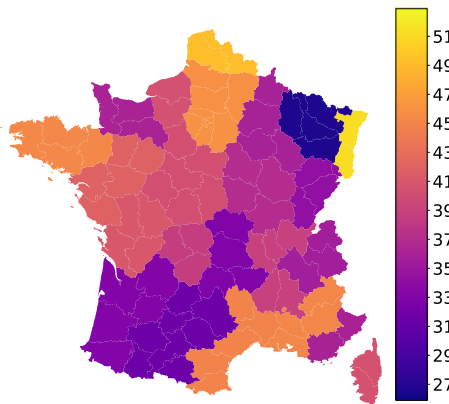


FIGURE 4 – Pourcentage de temps de parole attribué aux femmes dans les éditions régionales d’actualités du 19/20 de France 3 en 2016

Une étude de corrélation préliminaire a été réalisée en croisant le pourcentage de parole des femmes avec le nombre d’habitants par département. L’estimation de la corrélation a été réalisée en utilisant le test non-paramétrique de Spearman (Zwillinger & Kokoska, 2000). Une corrélation positive modérée ( $\rho = 0.453$ ) et statistiquement significative ( $pvalue < 10^{-5}$ ) tend à montrer que plus un département est peuplé, plus le taux de parole des femmes est élevé dans le 19/20.

## 5 Biais d'analyse

Les biais liés à cette approche automatique sont liés à la fiabilité du système de détection du genre du locuteur, qui n'a été entraîné et évalué que sur des voix d'adultes. Ces biais ont été réduits en excluant de l'analyse les chaînes thématiques ciblant les enfants, ainsi que les tranches horaires fréquemment associées à la diffusion de dessins animés (6-9h du matin). La prise en charge des voix d'enfants n'est pas aisée, et il existe assez peu de ressources audiovisuelles permettant d'entraîner et d'évaluer de tels systèmes (Schuller *et al.*, 2013). De plus, les voix d'enfants diffusées dans les dessins animés, dans les programmes doublés en français, ou dans la publicité radiophonique, sont très majoritairement interprétées par des acteurs professionnels adultes, n'ayant pas nécessairement le même sexe que le personnage qu'ils incarnent. Des observations informelles tendent à montrer que les « *voix d'enfant* » sont tantôt considérées comme de la musique (personnages de dessins animés très théâtraux), ou comme des voix de femme.

## 6 Conclusion

Cette étude présente un système d'analyse automatique de la parole fondé sur l'utilisation de réseaux de neurones convolutionnels capables de segmenter des archives audiovisuelles en zones de musique et en zones de paroles attribuées à des hommes ou à des femmes. Ce système a permis de décrire les différences de traitement existant dans les médias entre les hommes et les femmes, via l'estimation du *taux d'expression*. Il a été montré que sur la totalité des chaînes nationales, le temps de parole est attribué dans une plus grande proportion à des hommes, plus particulièrement dans les chaînes à contenu sportif ou culturel. Cette observation corrobore d'autres études concluant que *Le sérieux d'Arte se fait donc avec les hommes ; l'émotion de M6 se fait avec les femmes* (Reiser & Gresy, 2008). Une augmentation statistiquement significative du taux d'expression féminine a été observé sur 7 chaînes nationales entre 2010 et 2017, ce qui laisse supposer qu'une partie ces inégalités de traitement tend à se réduire.

Cette étude constitue à notre connaissance le premier cas d'usage de systèmes de reconnaissance automatique du genre du locuteur appliqué à la description du taux d'expression féminine. Il s'agit également de la description réalisée sur la plus grande quantité de données, correspondant à plus de 170 000 heures de flux télévisuels. Cette masse de données permet de réduire le biais associé au contexte de collecte des archives mentionné dans les analyses réalisées via des processus d'annotation manuelle (Reiser & Gresy, 2008; Levant *et al.*, 2014).

Les travaux en cours consistent à approfondir les analyses présentées, en y incluant des mesures d'audience par tranche horaire, ainsi qu'une sélection de stations radio (Doukhan *et al.*, 2018b). Des travaux de plus long terme consisteraient à décrire les différences de traitement des hommes et des femmes dans les médias à l'aide d'un plus grand nombre de descripteurs. L'utilisation de systèmes de transcription automatique de la parole devrait permettre d'estimer assez facilement les *taux d'identification*, et pourrait dans une moindre mesure aider à déterminer le statut des locuteurs (expert, témoin, etc...). L'estimation des *taux de présence* pourrait également être envisagée en se basant sur des procédures de segmentation et regroupement en locuteurs.



# Références

- CSA (2015). Les chiffres clés de l'audiovisuel français, édition du premier semestre.
- CSA (2017). La représentation des femmes à la télévision et à la radio - rapport sur l'exercice 2016.
- DOUKHAN D., CARRIVE J., VALLET F., LARCHER A. & MEIGNIER S. (2018a). An open-source speaker gender detection framework for monitoring gender equality. *Acoustics, Speech and Signal Processing (ICASSP)*.
- DOUKHAN D., POELS G. & CARRIVE J. (2018b). Describing gender equality in french audiovisual streams with a deep learning approach (accepted). *Journal of European Television History and Culture (VIEW)*.
- EL SHAFEY L., KHOURY E. & MARCEL S. (2014). Audio-visual gender recognition in uncontrolled environment using variability modeling techniques. In *International Joint Conference on Biometrics (IJCB)*, p. 1–8 : IEEE.
- EUROPE (2015). Toolkit » sur la mise en application de la recommandation du comité des ministres du conseil de l'europe cm/rec (2013) 1 sur l'égalité entre les femmes et les hommes et les médias. *Conseil de l'Europe, Commission pour l'égalité entre les femmes et les hommes*.
- GUADALUPE F. (2016). Audiences access : Le "19/20" de france 3 leader, "c à vous" et "28 minutes" en forme. <http://www.ozap.com>.
- LAMEL L. F. & GAUVAIN J.-L. (1995). A phone-based approach to non-linguistic speech feature identification. *Computer Speech & Language*, **9**(1), 87–103.
- LARCHER A., LEE K. A. & MEIGNIER S. (2016). An extensible speaker identification sidekit in python. In *Acoustics, Speech and Signal Processing (ICASSP)*, p. 5095–5099 : IEEE.
- LEVANT B., HANNOT M. & DERINOZ S. (2014). How gender representations matter with generation in television? In *II International Conference Gender and Communication*, p. 17–27.
- MACHARIA S. E. A. (2015). *Who Makes the News? : Global Media Monitoring Project 2015*. World Association for Christian Communication.
- MEFFRE B. (2017). Audiences access : Nagui leader en baisse, "le 19/20" devant "dna", "c à vous" en forme. <http://www.ozap.com>.
- OZAP (2010). France 3 : Le 19/20 au dessus des 20% de parts d'audience cette semaine. <http://www.ozap.com>.
- PÉPIOT E. (2015). Voice, speech and gender : male-female acoustic differences and cross-language variation in english and french speakers. *Corela. Cognition, représentation, langage*, (HS-16).
- REISER M. & GRESY B. (2008). L'image des femmes dans les médias. *Secrétariat d'Etat à la solidarité*.
- SALMON F. & VALLET F. (2014). An effortless way to create large-scale datasets for famous speakers. In *LREC*, p. 348–352.
- SCHULLER B., STEIDL S., BATLINER A., BURKHARDT F., DEVILLERS L., MÜLLER C. & NARAYANAN S. (2013). Paralinguistics in speech and language—state-of-the-art and the challenge. *Computer Speech & Language*, **27**(1), 4–39.
- XIA R., DENG J., SCHULLER B. & LIU Y. (2014). Modeling gender information for emotion recognition using denoising autoencoder. In *Acoustics, Speech and Signal Processing (ICASSP)*, p. 990–994 : IEEE.
- ZWILLINGER D. & KOKOSKA S. (2000). *CRC standard probability and statistics tables and formulae*. Chapman & Hall : New York.