CSA
CONSEIL SUPERIEUR DE L'AUDIOVISUEL

Note méthodologique

BAROMETRE

L'égalité dans les programmes
de radio en 2024 et 2025

Table des matières

Cette note présente en détail la méthodologie mise en œuvre dans le Baromètre de l'égalité dans les programmes radiophoniques de la Fédération Wallonie-Bruxelles (FWB) en 2024 et en 2025.

La méthodologie ayant été complètement repensée par rapport aux éditions précédentes afin d'intégrer l'utilisation d'outils recourant à l'intelligence artificielle (IA)[1], la présente note commence par exposer brièvement les questions que soulève l'intégration de tels outils dans une étude portant sur le genre, et plus particulièrement sur le genre que l'on peut attribuer aux voix.

La note détaille ensuite la méthodologie de chacun des deux volets de l'étude, le « Grand angle » et le « Focus », en présentant le corpus étudié, les unités d'encodage et de mesure mises en œuvre, la manière dont les données ont été traitées par les outils d'IA, la manière dont les résultats de ce traitement ont été soit évalués (dans le « Grand angle »), soit systématiquement vérifiés et enrichis par un encodage manuel (dans le « Focus ») et, enfin, l'analyse qui a été réalisée pour faire sens des résultats.

[1] Voir l'introduction de la synthèse ou du rapport complet de l'étude.

1. Intelligence artificielle et genre : bref cadrage préalable

L'utilisation d'outils de catégorisation automatique recourant à l'intelligence artificielle dans le cadre d'une étude portant sur le genre perçu des voix soulève différentes questions liées aux caractéristiques de ces outils, qui appellent à cadrer strictement leur utilisation dans le cadre de la présente étude.

La première de ces questions renvoie aux risques d'erreur et de biais des algorithmes de catégorisation, ainsi qu'au manque de transparence de ces algorithmes. Le premier aspect de cette question a déjà été largement discuté pour la reconnaissance automatique du genre sur la base d'images (Buolamwini et al., 2018 ; Schwemmer et al., 2020 ; Waelen et al., 2022), qui s'est avérée meilleure, pour de nombreux outils, pour les hommes que pour les femmes – l'écart se creusant plus encore entre les hommes blancs et les femmes noires.

Cette différence de traitement s'explique largement par un manque de diversité des données avec lesquelles les algorithmes sont entraînés : si les images fournies au modèle pour apprendre à reconnaître le genre d'une personne sont majoritairement des images d'hommes (blancs qui plus est), l'algorithme deviendra plus performant pour catégoriser les hommes (blancs en particulier). Il existe cependant différentes mesures permettant d'évaluer la performance d'un tel algorithme, et particulièrement les éventuels écarts de performance en fonction du genre des individus.

Cette problématique est intimement liée à une autre : l'absence de transparence des modèles de catégorisation automatique du genre. Le modèle apprenant par lui-même – c'est le principe même du *machine learning* – plutôt que sur la base d'un ensemble de règles qu'on lui fournirait au préalable, il est très difficile de savoir sur quels éléments il se fonde pour décider que telle image ou telle voix représente plutôt un homme ou plutôt une femme. C'est pourquoi on qualifie souvent ces algorithmes de « boîtes noires ».

L'absence de transparence des modèles apparaît comme un véritable enjeu, afin notamment de prévenir des erreurs et biais de reconnaissance. Il faut néanmoins souligner qu'il n'existe pas de définition stricte et absolue des voix féminines et masculines – au-delà du fait que la hauteur et le timbre de la voix sont des facteurs de distinction importants. La façon dont les êtres humains opèrent la catégorisation genrée des voix est donc, dans une certaine mesure, elle-même une « boîte noire ». Par ailleurs, le manque de transparence des algorithmes peut être en partie mitigé par des procédures d'évaluation des performances des modèles, permettant d'identifier et de quantifier les éventuels biais de leur fonctionnement – même si ce fonctionnement reste opaque.

Dans le cadre de cette étude, nous avons mis en place une procédure de vérification systématique des catégorisations automatiques du genre là où c'était possible et nous avons procédé à une évaluation des performances de la reconnaissance automatique du genre sur nos données lorsqu'une vérification manuelle systématique n'était pas possible. La présente note détaille la procédure d'évaluation mise en œuvre[2].

[2] Voir ci-dessous, les parties consacrées au « Traitement automatique par des modèles d'IA » et à la « Pré-annotation par des modèles d'IA ».

Une deuxième grande question soulevée par l'utilisation de modèles de reconnaissance automatique du genre concerne les possibilités mêmes de catégorisation de ces outils. En effet, même si certaines recherches commencent à se pencher sur la possibilité de catégoriser les voix sur un continuum, les outils disponibles au début de cette recherche se limitaient à une catégorisation binaire entre voix féminines et voix masculines. L'utilisation de tels modèle reviendrait, selon Foad Hamidi et ses collègues (2018), à exclure, invalider et assurer une catégorisation inexacte des personnes à l'identité de genre non-binaire. À cet égard, les chercheurs soulignent trois grandes préoccupations exprimées par un échantillon de personnes non-binaires :

- l'identité de genre ne peut pas se déduire des caractéristiques physiques d'une personne – et ce, que cette déduction soit faite par un être humain ou une machine ;
- l'utilisation de modèles de catégorisation automatique du genre représente une menace quant à l'autonomie et à la vie privée des individus ;
- en se limitant à une catégorisation binaire du genre, ces modèles risquent de renforcer les représentations stéréotypées du féminin et du masculin au sein de la société.

L'utilisation d'un modèle de catégorisation automatique du genre a été strictement cadrée, au sein de cette étude, pour prendre en compte ces préoccupations. Ainsi, comme on l'a déjà indiqué, l'étude ne prétend nullement déterminer l'identité de genre des personnes qui s'expriment à la radio, mais refléter les perceptions que les auditeurs et auditrices peuvent avoir à partir de l'indice de genre que représente la voix.

Dès lors, la question du respect de l'autonomie des personnes soumises au modèle ne concerne plus l'autodétermination de leur identité de genre, mais uniquement le fait d'être soumises à une catégorisation automatique. Il faut en outre rappeler que cette catégorisation est ici réalisée à des fins de description dans un cadre de recherche, et encadrée par une procédure de vérification et d'évaluation des performances du modèle d'IA. Les données obtenues grâce au modèle sont en outre traitées de manière agrégée et dans le respect des lois protégeant la vie privée des individus.

Enfin, la question de la catégorisation binaire nous semble particulièrement importante : si la mission qui nous a été confiée par le législateur concerne bien la représentation des femmes et des hommes, le CSA est, depuis longtemps, sensible à la question des identités de genre diverses. En outre, le risque de renforcer les représentations stéréotypées du féminin et du masculin au travers d'une catégorisation binaire va à l'encontre de l'objectif même de la présente étude.

À cet égard, il nous faut encore insister encore sur le fait que cette étude ne s'intéresse pas à l'identité de genre en tant que telle, mais à l'indice de genre que représente une voix pour qui l'entend. Dans une telle optique, la voix peut-elle représenter un indice de genre non-binaire interprétable pour le public de la radio ?

D'une part, les études qui se penchent sur les rapports entre voix et identités de genre diverses font apparaître une multiplicité de positionnements des personnes non-binaires par rapport à leur voix (Arnold, 2015 ; Candea et Brown, 2022) : volonté de masculinisation ou, à l'inverse, de féminisation, recherche de « neutralisation » ou encore détermination à ne pas altérer sa voix. Si la voix fait incontestablement partie des ressources mobilisables pour exprimer son identité de genre, il n'y a donc pas un type de voix spécifique qui pourrait correspondre à ces positionnements multiples – ce qui peut expliquer, selon ces mêmes études, les différences d'appréciation que l'on peut observer de la part des auditeurs et auditrices par rapport aux voix de personnes se définissant comme non-binaires.

D'autre part, malgré la visibilité croissante des identités de genre diverses, le schéma binaire reste très prégnant. Ainsi, lors d'une expérience leur demandant d'attribuer un genre à différentes voix – de femmes cisgenres, d'hommes cisgenres et de femmes transgenres –, très peu de participantes et de participants français.es ont choisi l'option « je ne sais pas » (5% pour les personnes non-binaires), pourtant disponible à côté des options « femme » et « homme » (Doukhan et al., 2024). Dans une autre expérience d'attribution d'un genre à partir de l'écoute de voix, utilisant cette fois une échelle continue en 15 points entre le féminin et le masculin, la zone médiane de l'échelle (soit le tiers situé entre la zone féminine et la zone masculine) a été utilisée par environ 25% des participantes et participants français.es pour évaluer les voix de

personnes non-binaires (Candea et Brown, 2022). Si on peut relever que cette proportion est bien plus élevée que dans l'expérience précédente, les chercheuses ne précisent cependant pas comment ces évaluations se répartissent au sein de la zone, autour du point médian. En outre, il n'est pas possible de vérifier si l'ensemble de la zone médiane a effectivement été perçu comme « neutre » par les participantes et les participants.

Enfin, dans leur étude de l'influence des stéréotypes de genre sur la perception des voix, Aron Arnold et Maria Candea (2015) ont notamment testé la façon dont sont perçues des voix resynthétisées de manière à s'écarter des prototypes de voix féminine et masculine. Dans ce cadre, les chercheurs ont considéré comme « androgynes » les voix resynthétisées qui ont reçu autant d'évaluations comme voix féminines que comme voix masculines de la part d'un panel composé de dix personnes. Lors de l'analyse des résultats de leur expérience, les chercheurs font l'hypothèse que ces voix androgynes ont été évaluées par rapport aux prototypes de voix féminine et masculine : « Les voix présentées comme voix de femmes, perçues comme graves et sombres par rapport aux prototypes, ont pu évoquer des attitudes associées cognitivement aux basses fréquences [et donc aux voix masculines], tandis que les voix présentées comme voix d'hommes ont pu évoquer des attitudes associées aux fréquences élevées [et donc aux voix féminines] » (Arnold et Candea, 2015, p. 84).

Dans un tel contexte, il apparaît difficile de considérer que la voix peut constituer, à l'heure actuelle, un indice de genre univoque renvoyant à des identités de genre non-binaires – même si cette situation est bien sûr susceptible d'évoluer, parallèlement à la reconnaissance et à l'inclusion des identités de genre diverses au sein de la société (Candea et Brown, 2022 ; Brown et Pillot-Loiseau, 2022).

Pour les catégorisations automatiques du genre, réalisées entièrement par des outils recourant à l'IA, on parlera dans la présente étude de voix féminines et de voix masculines – sans présumer de l'identité de genre des locuteurs et locutrices. Pour les catégorisations vérifiées manuellement, on parlera d'hommes, de femmes et de personnes non-binaires en croisant l'indice de la voix avec l'ensemble des indices de genre traversant les discours des personnes qui parlent – tout en restant conscient qu'il s'agit bien là de la perception des personnes ayant réalisé l'encodage.

En outre, dans le souci de continuer à prendre en compte la question des identités de genre diverses, l'étude présente une analyse de la présence et de l'utilisation d'un lexique de termes renvoyant à ces identités dans la transcription automatique des programmes composant le corpus de l'étude. Cette analyse permet d'évaluer dans quelle mesure et de quelle façon la question des identités de genre diverses est abordée à la radio en FWB[3].

[3] Voir ci-dessous, la partie consacrée à l'« Analyse de la transcription des programmes » dans le « Grand angle ».

Grand angle

Estimation automatique du temps de parole des voix féminines et masculines sur l'ensemble des programmes

2. Grand angle

Le premier volet de l'étude, intitulé « Grand angle », consiste en une estimation automatique, réalisée grâce à des modèles d'intelligence artificielle, de la présence des voix féminines et masculines dans les programmes de 27 services radiophoniques diffusés en FM en FWB, sur une durée de 7 jours en 2024 et 7 jours en 2025.

2.1. Corpus

Le volet « Grand angle » de l'étude porte sur deux semaines reconstituées, soit des semaines dont les jours ont été sélectionnés au hasard :

- entre la rentrée scolaire et la fin de l'année 2024 (lundi 16 décembre, mardi 15 octobre, mercredi 11 septembre, jeudi 24 octobre, vendredi 30 août, samedi 2 novembre et dimanche 24 novembre) ;
- entre la rentrée et la fin de l'année 2025 (lundi 17 novembre, mardi 30 septembre, mercredi 3 décembre, jeudi 4 septembre, vendredi 10 octobre, samedi 18 octobre et dimanche 9 novembre)[4].

La méthode d'échantillonnage par semaine reconstituée vise à limiter l'impact d'un événement particulier sur le corpus d'analyse tout en tenant compte du caractère cyclique de la temporalité des médias. Le choix de la période allant de la rentrée scolaire à la fin de l'année, en excluant la période des fêtes de fin d'année, vise quant à lui à assurer la cohérence des grilles de programmes pour l'ensemble des jours sélectionnés.

Le corpus est composé des 27 services radiophoniques FM dont l'audience potentielle est la plus importante[5] et dont au moins 75% des programmes sont en langue française[6] :

- 5 radios publiques de la RTBF : Classic 21, La Première, Musiq3, Tipik, VivaCité (Bruxelles) ;
- 6 radios privées en réseau à couverture communautaire (couvrant l'ensemble du territoire de la FWB) ou urbaine (couvrant les centres urbains et les axes routiers principaux de la FWB) : Bel RTL, Fun Radio, LN Radio, Nostalgie, NRJ, Radio Contact ;
- 4 radios privées en réseau à couverture provinciale : Antipode, Inside Radio, Maximum FM, Sud Radio ;
- 5 radios privées indépendantes : Arabel, BX FM, K.I.F, Radio Judaïca, Vibration ;
- 7 radios privées indépendantes ayant le statut de « radios associatives et d'expression à vocation culturelle ou d'éducation permanente »[7] : 48 FM, Equinoxe FM, Radio Air Libre, Radio Campus Bruxelles, RCF Bruxelles, RCF Liège, Radio Panik.

Le corpus de ce volet « Grand angle » comprend l'intégralité de la programmation diffusée sur ces 27 chaînes tout au long des 7 jours composant chacun des deux échantillons étudiés[8]. Il est ainsi composé de 4.536 heures de flux audio pour chaque année analysée, soit 9.072 heures au total[9].

[4] A la suite d'un problème d'enregistrement, le dimanche 9 novembre a dû être remplacé par le 2 novembre pour Classic 21.
[5] L'audience potentielle d'une radio étant le nombre de personnes résidant dans la zone de couverture de la radio et qui sont donc, potentiellement, en mesure de l'écouter. Le seuil a été fixé à 5% de la population de la FWB. 28 radios s'avèrent être au-dessous de ce seuil, mais Warm FM a dû être écartée du corpus après le traitement automatique par les outils d'IA car la part d'émissions s'y est avérée trop faible pour que les résultats de l'estimation automatique restent fiables.
[6] Selon les résultats du contrôle effectué par le CSA pour les années 2022-2023. Ce critère lié à la langue a été introduit car les modèles d'IA utilisés se basent notamment sur la transcription des émissions.
[7] Ce statut est délivré par le Collège d'autorisation et de contrôle du CSA aux radios indépendantes qui ont recours principalement au volontariat et qui consacrent l'essentiel de leur programmation à des émissions d'information, d'éducation permanente, de développement culturel et de participation citoyenne. Les radios bénéficiant de ce statut reçoivent chaque année une subvention de fonctionnement.
[8] En ce compris les programmes qui ne relèveraient pas de la production propre de l'éditeur et les éventuelles rediffusions. Alors que les Baromètres précédents consacrés aux programmes télévisés ne portaient que sur les programmes relevant de la production propre des éditeurs, le Baromètre des programmes radios de 2019 introduisait une exception à ce principe afin « d'avoir un aperçu de la manière dont les matinales des chaînes les plus écoutées en Fédération Wallonie-Bruxelles représentent l'égalité » (CSA, 2019, p. 9). Cette nouvelle édition du Baromètre poursuit cette logique en l'étendant à l'ensemble des programmes des chaînes les plus importantes de la FWB.
[9] En raison de problèmes techniques lors de l'enregistrement, quelques fichiers n'ont cependant pas pu être traités ou n'ont pas pu être traités dans leur entièreté. La liste complète des fichiers concernés est disponible en annexe.

2.2. Unités d'encodage et de mesure

Alors que les Baromètres précédents visaient à quantifier l'ensemble des personnes intervenant dans les médias audiovisuels[10] (celles qui parlent, dont on parle, que l'on voit), le volet « Grand angle » de cette nouvelle étude s'intéresse à une catégorie particulière des personnes intervenant au sein des programmes radiophoniques : les personnes qui prennent la parole[11]. Ce choix d'unité d'encodage, plus restreint que dans les éditions précédentes, correspond aux intervenant.es les plus directement perceptibles pour le public. Il s'agit en outre de la catégorie de personnes intervenant dans les émissions la plus importante en radio. Ainsi, dans le Baromètre des programmes radiophoniques de 2019, les personnes prenant la parole représentaient 52,89% de l'ensemble des personnes encodées.

Ce recentrage sur les personnes qui prennent la parole s'accompagne d'un changement d'unité de mesure. Les Baromètres précédents proposaient un décompte du nombre de personnes intervenant dans les émissions, dans lequel chaque personne avait le même poids, qu'elle présente toute une émission ou qu'elle apparaisse quelques secondes à l'arrière-plan de l'image. Ce « Grand angle » entend quant à lui quantifier le temps de parole et, ainsi, appréhender plus précisément la place respective accordée aux voix féminines et aux voix masculines dans les programmes.

Ces choix s'inscrivent cependant dans une forme de continuité par rapport aux précédentes éditions du Baromètres : celles-ci ont permis « de dresser un état des lieux brut » et « de comparer son évolution au fil du temps » (CSA, 2021, p. 8). Elles constituent un préalable essentiel à la présente étude en documentant la situation, et ses faibles évolutions, pendant 10 ans.

2.3. Traitement automatique par des modèles d'IA

Le corpus du volet « Grand angle » de cette étude a été traité par trois modèles d'intelligence artificielle permettant d'estimer automatiquement le temps de parole des voix féminines et masculines dans les programmes diffusés par les 27 chaînes durant les deux semaines reconstituées composant les échantillons[12]. Chaque modèle d'IA permet d'annoter les flux audio qui lui sont soumis au travers de deux opérations :

- Une opération de découpage, qui consiste à délimiter différents segments au sein du flux (en précisant le moment exact du début et de la fin de chaque segment) ;
- Une opération de classification, au cours de laquelle chaque segment est associé à une catégorie spécifique (en fonction du modèle d'IA utilisé).

Dans le volet « Grand angle » de l'étude, les annotations ont été entièrement effectuées par les modèles d'IA, sans vérification humaine. C'est pourquoi on parle d'« estimation automatique ». Lorsque l'on interprète les résultats de cette estimation automatique, il est essentiel de garder en tête les métriques d'évaluation des performances des différents modèles, ainsi que les biais déjà documentés[13].

Identification des programmes, de la publicité et de la musique

Un premier modèle d'intelligence artificielle permet de distinguer, au sein des flux radiophoniques qui lui sont soumis, ce qui relève des programmes/émissions[14], de la publicité et de la musique. Ce modèle a été développé en étroite collaboration avec notre partenaire pour ce projet, Multitel, spécifiquement pour la présente étude.

[10] Dans les médias télévisuels ou radiophoniques, au sein des programmes ou de la communication commerciale, selon les éditions.
[11] L'étude ne s'intéresse donc pas aux voix que l'on entend dans les spots publicitaires, ni aux interprètes des titres musicaux, ni aux personnes dont on parle mais qui ne prennent pas la parole et aux personnes que l'on entend en arrière-fond dans les programmes.
[12] La présente note n'ayant pas pour vocation de documenter de manière technique la façon dont fonctionne l'application développée par Multitel dans le cadre de cette étude, elle ne décrit pas l'ensemble des tâches réalisées par cette application. Dès lors, l'ordre dans lequel les modèles d'IA sont présentés ici ne correspond pas forcément à l'ordre dans lequel l'application fait intervenir ces différents modèles.
[13] Voir ci-dessous, le bref cadrage préalable portant sur « Intelligence artificielle et genre ».
[14] On utilisera indifféremment l'un ou l'autre de ces termes dans cette note.

Le modèle devait non seulement être capable de détecter la présence de parole, mais également de comprendre les caractéristiques contextuelles et acoustiques permettant de différencier chaque type de segment. Par exemple, les publicités contiennent généralement une parole très structurée et dynamique, souvent accompagnée de musique de fond ou d'effets sonores, tandis que les segments musicaux présentent des propriétés rythmiques et harmoniques associées aux chansons. Les segments d'émissions, quant à eux, se caractérisent davantage par une parole conversationnelle naturelle, des interactions entre plusieurs personnes et des structures acoustiques moins répétitives. Le modèle devait être capable d'identifier ces différences subtiles avec suffisamment de précision et de robustesse pour être exploitable dans un environnement réel de diffusion radio.

Le modèle développé par Multitel repose sur une architecture hybride de *deep learning* qui combine un réseau de neurones convolutionnel (ou CNN pour « Convolutional Neural Network ») avec un réseau de mémoire à long terme bidirectionnelle (ou BiLSTM pour « Bidirectional Long Short-Term Memory »), particulièrement efficace pour réaliser des tâches complexes impliquant des données séquentielles (comme le son), où les caractéristiques locales et la compréhension du contexte sont toutes deux essentielles (comme pour distinguer une émission d'une publicité).

Le modèle a été entraîné sur un ensemble de 174 fichiers audio d'une durée d'une heure, annotés manuellement, ce qui a permis au modèle d'apprendre les caractéristiques distinctives de chaque catégorie de segment et d'améliorer ses performances de prédiction.

Ces performances ont ensuite été évaluées sur un sous-échantillon du corpus du « Grand angle », constitué de 220 fichiers audio d'une heure ayant fait l'objet d'un encodage manuel délimitant les segments relevant des émissions, des publicités et de la musique.

Les métriques d'évaluation qui ont été utilisées dans le cadre de cette évaluation sont :

- la précision, qui calcule, parmi tous les segments audio d'une seconde classés dans une catégorie par le modèle, la proportion qui relève effectivement de cette catégorie ;
- le rappel (ou la sensibilité), qui mesure, parmi tous les segments audio d'une seconde qui appartiennent réellement à une catégorie, la proportion de segments correctement classés par le modèle ;
- le F1-score, qui constitue une moyenne harmonique entre la précision et le rappel, donnant une balance entre les deux ;
- la matrice de confusion, un tableau permettant de visualiser où le modèle fait erreur.

Le tableau suivant présente les résultats obtenus pour les trois premières métriques

Si le modèle obtient assez logiquement de meilleures métriques pour la catégorie « musique », ses performances pour identifier les émissions restent excellentes, avec un F1-score de 95,85%.

Catégorie	Précision	Rappel	F1-score
Emission	94,88%	96,85%	95,85%
Publicité	88,47%	86,46%	87,45%
Musique	96,90%	98,88%	97,88%

Tableau 1

La matrice de confusion, ci-dessous, permet de croiser les catégories réelles (sur l'axe vertical) et les prédictions du modèle (sur l'axe horizontal) pour l'ensemble des segments audio ayant servis à l'évaluation[15].

La matrice de confusion indique ici que :

- 1% des segments d'une seconde que le modèle classe comme des publicités constituent en fait des émissions ;
- 2% des segments d'une seconde que le modèle classe comme de la musique constituent en fait des émissions ; 1% des segments d'une seconde que le modèle classe comme des émissions constituent en fait de la musique ;
- 12% des segments d'une seconde que le modèle classe comme des émissions constituent en fait des publicités.

L'erreur la plus fréquente du modèle est donc de considérer des publicités comme des émissions.

Afin d'améliorer encore les performances du modèle, la détection des publicités a été renforcée au travers de la comparaison des flux audio traités avec une base de données constituée des empreintes sonores de publicités ayant fait l'objet d'annotations manuelles. Ainsi, si une publicité de la base de données est diffusée dans un nouveau fichier audio, elle est automatiquement reconnue en tant que telle.

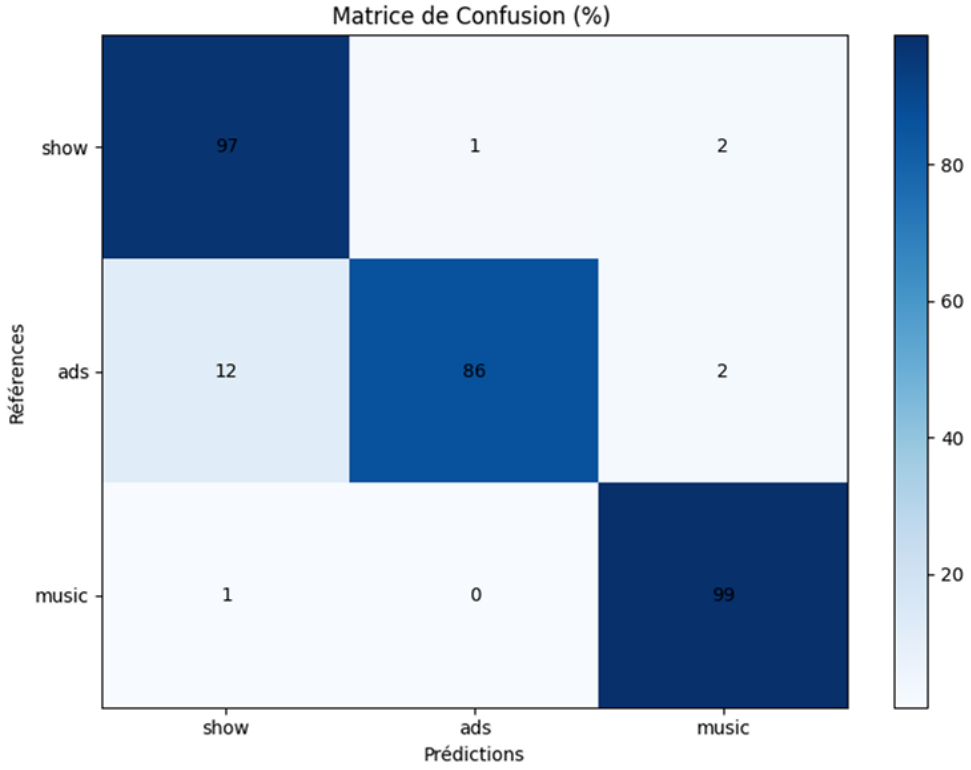

Figure 1

[15] Plus la couleur d'une case de la matrice est foncée, plus la proportion de cas relevant de cette case est élevée. On espère donc, lorsqu'on réalise une matrice de confusion, que les cases croisant les mêmes catégories sur les deux axes soient aussi foncées que possible – et les autres cases aussi claires que possible.

Reconnaissance du genre

Pour tous les segments audio identifiés automatiquement comme programmes, un deuxième modèle d'IA permet ensuite de catégoriser toutes les voix contenues dans ces segments comme voix féminine ou voix masculine. Cette tâche est réalisée par le modèle en libre accès Ecapa.

Ecapa est un modèle conçu pour reconnaître et différencier les voix humaines. Son objectif principal consiste à transformer un signal audio (tel qu'un enregistrement de parole) en une représentation numérique (ou vecteur de voix, pour « voice embedding ») qui capture les caractéristiques uniques de la voix d'une personne : hauteur, formants, timbre, dynamique. Le genre se manifeste fortement dans ces indices, ce qui signifie que le vecteur de voix contient les informations nécessaires à un classificateur fiable du genre. Le modèle a été entraîné sur de vastes jeux de données diversifiés et est spécialement conçu pour gérer l'audio du monde réel (réverbération, bavardage de fond et une variété d'équipements d'enregistrement) afin que les prédictions de genre restent précises même dans les environnements acoustiques les plus chaotiques.

Le modèle Ecapa a été choisi en raison de ses performances déjà documentées. Cependant, étant donné son rôle central dans le volet « Grand angle » de la présente étude, une procédure visant à évaluer ses performances sur nos données a été mise en place par Multitel.

L'évaluation a été réalisée sur les émissions contenues dans 110 fichiers audio d'une heure ayant fait l'objet d'annotations manuelles tant pour identifier les émissions que le genre des personnes y prenant la parole.

Le tableau qui suit présente les résultats obtenus lors de l'évaluation en termes de précision, rappel et F1-score.

Les performances du modèle sont donc proches pour les deux catégories, bien que légèrement supérieures pour les voix masculines, avec un F1-score de 93,89%, que pour les voix féminines, avec un F1-score de 92,51%.

Si ces scores sont légèrement en-dessous des évaluations réalisées sur d'autres bases de données, ils restent très bons pour ce type d'outil. La différence de performance s'explique principalement par le caractère complexe des flux radiophoniques, qui présentent notamment beaucoup de changements rapides de locuteurs et de locutrices, des niveaux de qualité sonore variables pour les interventions d'auditeurs et d'auditrices ou les interviews hors studio, l'utilisation de fonds sonores en arrière-plan des prises de parole, etc.

En outre, l'évaluation sur nos données a été réalisée de la manière la plus « sévère » possible, en considérant comme des erreurs de reconnaissance, par exemple, les silences entre deux prises de parole d'une même personne.

Catégorie	Précision	Rappel	F1-score
Voix féminines	90,32%	94,80%	92,51%
Voix masculines	94,16%	93,63%	93,89%

Tableau 3

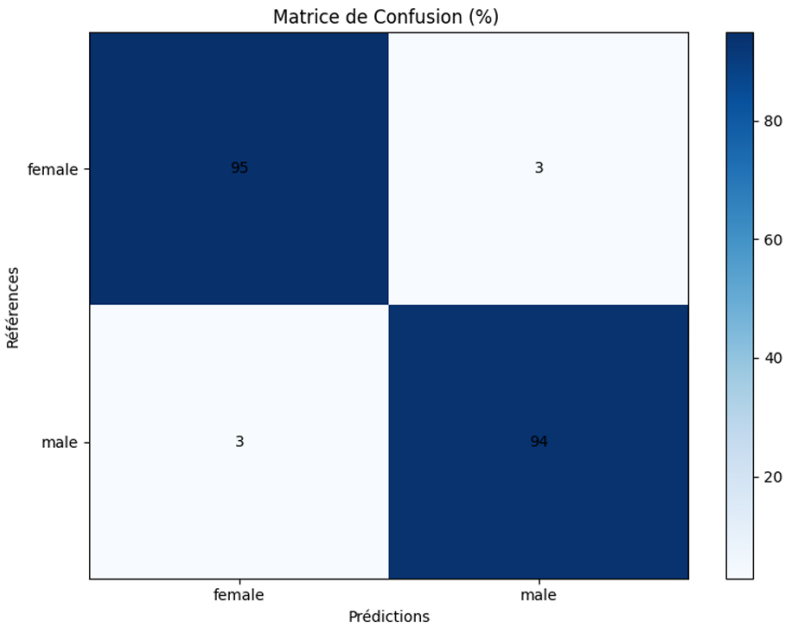

Figure 1

Si on se concentre sur les erreurs qui ne sont pas liées à des silences, comme dans la matrice de confusion ci-dessus, on observe que :

- 3% des segments d'une seconde que le modèle classe comme des voix féminines sont en fait des voix masculines ;
- 3% des segments d'une seconde que le modèle classe comme des voix masculines sont en fait des voix féminines.

Transcription

Enfin, un troisième modèle d'IA a été utilisé pour transcrire automatiquement tous les flux audio analysés sous la forme de texte. Cette transcription est nécessaire pour permettre l'analyse des propos tenus dans les programmes sous l'angle des identités de genre diverses, mais elle est aussi utilisée dans d'autres phases du traitement automatique des flux audio nécessitant une base textuelle. Le modèle utilisé pour la transcription est Whisper, disponible en source ouverte, de la société OpenAI, dont le taux d'erreur sur les mots (ou WER pour « Word Error Rate ») est évalué, dans la littérature scientifique, à 9,7%[16].

Plus précisément, on a utilisé ici la version Large V3 du modèle. L'utilisation d'un modèle « à grande capacité » (ou « large model ») est nécessaire dans des conditions où la parole peut varier en rythme, clarté, profil de locuteur et contexte acoustique, comme c'est le cas en radio. La version Large V3 affiche un taux d'erreur sur les mots (WER) significativement plus bas que les versions précédentes, surtout dans le cas de discours bruyants ou accentués.

[16] Etant donné le rôle plus périphérique de la transcription automatique dans l'étude et les performances déjà bien documentées du modèle utilisé, Whisper n'a pas fait l'objet d'une évaluation sur nos données. Voir Bain et al. (2023) pour les détails de l'évaluation.

2.4. Analyse des résultats de l'encodage automatique

La première étape du traitement des données par les modèles d'IA consiste en la catégorisation automatique des flux audio entre musique, publicité et programmes – ce qui permet ensuite de ne conserver que les segments considérés comme des programmes pour l'analyse des voix féminines et masculines.

Les résultats de cette catégorisation méritent cependant aussi d'être analysés en eux-mêmes : d'une part, il s'agit de données inédites qui permettent de mieux connaître les radios du corpus et, d'autre part, ces données permettent aussi d'affiner l'analyse de la présence des voix féminines et masculines au sein des programmes en tenant compte des différences qui existent dans les grilles de programmes en fonction du moment de diffusion.

Ensuite, le cœur de l'analyse porte sur les prises de parole des voix féminines et masculines au sein des programmes – à partir, toujours, de l'encodage automatique réalisé par les modèles d'IA. Les résultats de cette analyse sont présentés :

- en termes « absolus », au travers du temps de parole, qui permet d'appréhender les données en termes de volumes horaires et de comprendre comment ces volumes varient en fonction du moment de diffusion et selon les caractéristiques des radios analysées ;
- en termes « relatifs », au travers du taux de parole, soit le pourcentage de temps de parole des voix féminines par rapport à celui des voix masculines, qui permet des comparaisons plus aisées.

Le temps de parole et le taux de parole sont d'abord calculés de manière globale, sur l'ensemble des programmes, puis de manière plus fine en fonction du contexte de diffusion et des caractéristiques des radios reprises dans le corpus.

Il est important de tenir compte du jour et de la tranche horaire de diffusion car la part de temps d'antenne consacrée aux émissions est très variable entre la semaine et le week-end, ainsi que selon le moment de la journée. Ces différences de programmation sont d'autant plus importantes qu'elles sont liées à des différences en matière d'audience : c'est globalement aux moments où l'audience est la plus forte que la part d'émissions est la plus élevée. L'analyse sera donc affinée en fonction du jour et de la tranche horaire de diffusion.

Par ailleurs, les radios qui composent le corpus présentent des profils très diversifiés, et ce à plusieurs égards. L'analyse sera dès lors affinée selon :

1) **Le type et sous-type** : il peut s'agir de radios indépendantes (avec ou sans le statut de radios associatives) ou de radios en réseau (publiques, privées à couverture communautaire ou urbaine, privées à couverture provinciale).
2) **Le format :** ces radios présentent des projets éditoriaux et musicaux très diversifiés[17].

   Parmi les radios indépendantes :

   - les radios communautaires, dont tout le projet éditorial et musical est défini par la volonté de s'adresser à une communauté définie par un trait culturel particulier ;
   - les radios d'expression, dont le projet est orienté par des valeurs de participation citoyenne et de diversité culturelle ;
   - les radios thématiques, qui peuvent présenter des projets très divers mais dont le point commun est d'être centré sur un type particulier de contenus et/ou de style musical. Les trois radios thématiques du

[17] Lors de l'attribution des fréquences, une analyse des dossiers de candidature des différents services est réalisée par le CSA en vue de rattacher chaque projet à un format primaire, éventuellement un format secondaire et, dans le cas des radios de format généraliste, un sous-format – le Collège d'autorisation et de contrôle étant chargé d'« assurer une diversité du paysage radiophonique et un équilibre entre les différents formats de radios, à travers l'offre musicale, culturelle et d'information » (Décret SMA-SPV, Art. 3.1.3-4. - § 1er). Pour établir la typologie utilisée ici, nous avons utilisé le niveau de format qui permet de cerner au mieux les particularités de différentes chaînes du corpus, ainsi que leurs différences les unes par rapport aux autres.

corpus sont centrées sur la programmation de musiques électroniques (Vibration), un registre musical mêlant hip hop, rap et RnB (K.I.F) et des contenus sur l'entrepreneriat à Bruxelles et en Europe (BXFM).

Parmi les radios en réseau du corpus :

- les radios de news/talk, dont la programmation accorde une place importante aux émissions ;
- les radios musicales adultes, dont le projet est centré autour de la programmation de musique grand public à destination d'une cible adulte[18] ;
- les radios musicales jeunes, qui se caractérisent par une programmation musicale grand public destinée à un public jeune ;
- les radios géographiques, dont le projet est construit autour de leur zone de couverture géographique. Il s'agit ici uniquement des réseaux privés à couverture provinciale, qui présentent toutes une programmation musicale grand public importante et une faible part d'émissions.

3) **La part d'émissions :** il s'agit de la part de temps d'antenne consacrée aux programmes (et non à la musique et à la publicité).Les 27 radios ont été rassemblées en trois groupes selon la moyenne de la part de temps d'antenne qu'elles consacrent aux émissions en 2024 et 2025[19] : moins de 12,5%, entre 12,5% et 25% ou plus de 25%.

4) **L'audience :** cette partie de l'analyse se concentre sur les radios en réseau pour lesquelles on dispose de mesures d'audience fournies par le CIM[20]. Ces 15 radios ont été classées dans trois groupes, contenant 5 radios chacun[21] :
   - faible audience : de 0,07% à 0,70% ;
   - audience moyenne : de 1,51% à 6,97% ;
   - forte audience : de 7,71% à 10,58%.

Ces différentes particularités sont prises en compte dans les analyses développées dans ce « Grand angle ». Le tableau ci-dessous présente les caractéristiques de chacune des 27 radios étudiées – sans entrer dans le détail, chacune de ces caractéristiques faisant l'objet d'une explication au début des parties d'analyses correspondantes, ainsi que dans la note méthodologique.

[18] On a regroupé ici le format des radios musicales adultes à proprement parler et le format dit « patrimoine / back-catalog » qui s'adresse également au grand public adulte, mais au travers d'une programmation issue plus spécifiquement du patrimoine musical et moins orienté vers les nouveautés.

[19] Celle-ci a été calculée en faisant la moyenne de la proportion du temps d'antenne consacrée par chaque chaîne aux émissions en 2024 et en 2025 – cette proportion ayant été calculée à partir des résultats de la catégorisation automatique du temps d'antenne réalisée par l'outil d'IA entre musique, publicité et émission.

[20] Parmi les radios indépendantes du corpus, BXFM est la seule pour laquelle on dispose d'une mesure d'audience. Par souci de cohérence, elle n'est pas reprise dans l'analyse qui suit.

[21] Les chiffres d'audience utilisés pour définir les groupes sont une moyenne de l'audience moyenne quotidienne calculée par le CIM lors du dernier quadrimestre 2024 et de celle du dernier quadrimestre 2025 (soit les périodes dont notre corpus est tiré) – sauf pour Antipode pour laquelle on ne dispose que de l'audience moyenne quotidienne sur l'année 2024.

Service	Variables				
	Type	Sous-type	Format	Part d'émissions	Audience
48FM	Indépendante	Associative	Expression	Moins de 12,5%	N/A
Antipode	Réseau	Provincial	Géographique	Moins de 12,5%	Faible
Arabel	Indépendante	Non associative	Communautaire	De 12,5 et 25%	N/A
Bel RTL	Réseau	Communautaire/Urbain	News/Talk	25% et plus	Forte
BXFM	Indépendante	Non associative	Thématique	Moins de 12,5%	N/A
Classic 21	Réseau	RTBF	Musicale adulte	De 12,5 et 25%	Forte
Equinoxe FM	Indépendante	Associative	Expression	Moins de 12,5%	N/A
Fun Radio	Réseau	Communautaire/Urbain	Musicale jeune	De 12,5 et 25%	Moyenne
Inside Radio	Réseau	Provincial	Géographique	Moins de 12,5%	Faible
K.I.F	Indépendante	Non associative	Thématique	Moins de 12,5%	N/A
La Première	Réseau	RTBF	News/Talk	25% et plus	Moyenne
LN Radio	Réseau	Communautaire/Urbain	Musicale adulte	Moins de 12,5%	Faible
Maximum FM	Réseau	Provincial	Géographique	Moins de 12,5%	Faible
Musiq3	Réseau	RTBF	Musicale adulte	Moins de 12,5%	Moyenne
Nostalgie	Réseau	Communautaire/Urbain	Musicale adulte	De 12,5 et 25%	Forte
NRJ	Réseau	Communautaire/Urbain	Musicale jeune	De 12,5 et 25%	Moyenne
Radio Air Libre	Indépendante	Associative	Expression	De 12,5 et 25%	N/A
Radio Campus	Indépendante	Associative	Expression	De 12,5 et 25%	N/A
Radio Contact	Réseau	Communautaire/Urbain	Musicale adulte	De 12,5 et 25%	Forte
Radio Judaïca	Indépendante	Non associative	Communautaire	De 12,5 et 25%	N/A
Radio Panik	Indépendante	Associative	Expression	De 12,5 et 25%	N/A
RCF Bruxelles	Indépendante	Associative	Communautaire	25% et plus	N/A
RCF Liège	Indépendante	Associative	Communautaire	25% et plus	N/A
Sud Radio	Réseau	Provincial	Géographique	Moins de 12,5%	Faible
Tipik	Réseau	RTBF	Musicale jeune	De 12,5 et 25%	Moyenne
Vibration	Indépendante	Non associative	Thématique	Moins de 12,5%	N/A
VivaCité	Réseau	RTBF	News/Talk	25% et plus	Forte

Tableau 1

2.5. Analyse de la transcription des programmes

Si le modèle de reconnaissance automatique du genre utilisé dans cette étude ne permet qu'une catégorisation binaire des voix – féminines ou masculines –, le CSA reste attentif à la question des identités de genre diverses. C'est dans cette optique que le « Grand angle » propose une analyse de la présence et de l'utilisation d'un lexique de termes renvoyant à ces identités dans la transcription automatique du corpus effectuée par le modèle d'IA Whisper (*cf. supra*).

Le lexique a été construit en compilant plusieurs ressources lexicales relatives au genre de manière large[22], puis en sélectionnant les termes propres aux identités de genres et en excluant les termes présentant un fort risque de faux positif (comme « transition » ou « autodétermination ») ainsi que les termes les moins usités (tels qui « bispirituel » ou « AFAB »). Par ailleurs, certains termes figurent dans le lexique alors qu'ils ne sont pas acceptés par les communautés concernées ou que leur utilisation y fait l'objet de débats. Leur inclusion s'explique uniquement par le fait qu'ils restent largement utilisés en dehors de ces communautés – et ne représente en aucun cas une prise de position du CSA quant à leur utilisation.

La recherche au sein de la transcription fonctionnant sur la base de chaînes de caractères exactes, certains termes ont été réduits à leur partie invariable[23] tandis que d'autres ont été dédoublés pour tenir compte des différentes orthographes possibles[24]

28 termes du lexique relatif aux identités de genre diverses
Agenre ; Assignation à la naissance ; Changement de sexe ; Cisgenre ; Dysphorie de genre ; Expression de genre ; Gender-bending ; Genderfluid ; Genre fluide ; Hermaphrodite ; Identité de genre ; Identité sexuelle ; Intergenre ; Intersex* ; Mégenr* ; Non binaire;Non-binaire ; Non-binarité ; Non genré ; Queer ; Sexe assigné à la naissance ; Transexualité ; Transgenre ; Transidentité ; Transphob* ; Transsexuel ; Trouble de l'identité de genre ; Xénogenre

Tableau 2

Une vérification systématique des résultats de la recherche a été réalisée afin de :

- s'assurer qu'il ne s'agissait pas d'erreurs de transcription ;
- vérifier si le terme était utilisé dans un programme, dans un spot publicitaire ou dans un titre musical ;
- comprendre le contexte dans lequel chaque terme était utilisé.

L'analyse s'est ensuite attachée à :

- quantifier l'utilisation des différents termes ;
- de déterminer quand et où ils étaient utilisés ;
- qualifier le contexte dans lequel ils étaient utilisés.

[22] Les ressources en question provenaient de l'Institut pour l'égalité des femmes et des hommes (IEFH), de la Direction de l'égalité des chances de la FWB, de la campagne « Et toi t'es casé ? », de l'Association des journalistes professionnels (AJP), des associations *Genres pluriels* et *Equitas*, et du Mouvement français pour le Planning familial. Pour les références complètes, voir la bibliographie.
[23] Il s'agit de tous les termes se terminant par un astérisque dans le tableau qui suit.
[24] Sont ici concernées toutes les expressions formées par plusieurs mots, comme « expression de genre » pour lequel on a aussi recherché « expressions de genre ».

Focus

Analyse vérifiée manuellement du genre en termes de nombre de personnes et de temps de parole dans les émissions matinales

3. Focus

Le deuxième volet de l'étude, intitulé « Focus », consiste en une analyse, préparée par des outils d'IA puis entièrement vérifiée manuellement, de la place des femmes, des hommes et des personnes non-binaires au sein des émissions matinales des 11 radios FM en réseau à couverture communautaire et urbaine de la FWB, sur 3 jours de semaine et 1 jour de week-end en 2024.

3.1. Corpus

Le volet « Focus » de l'étude porte sur les 11 radios FM à couverture communautaire et urbaine de la FWB et sur 4 jours choisis de manière aléatoire au sein de l'échantillon du « Grand angle » de 2024 : lundi 16 décembre, mardi 15 octobre, vendredi 30 août et samedi 2 novembre[25].

Les plages horaires étudiées correspondent aux heures de grande audience radiophonique, soit la tranche allant de 6h à 10h en semaine et de 10h à 14h le week-end[26]. Tous les programmes diffusés dans ces tranches horaires ont été intégrés au corpus, y compris les éventuelles rediffusions.

Sont reprises dans ce deuxième volet de l'étude les 11 radios FM à couverture communautaire et urbaine de la FWB :

- 5 radios publiques de la RTBF : Classic 21, La Première, Musiq3, Tipik, VivaCité (Bruxelles) ;
- 6 radios privées en réseau à couverture communautaire ou urbaine : Bel RTL, Fun Radio, LN Radio, Nostalgie, NRJ, Radio Contact.

Il s'agit des mêmes plages horaires et des mêmes radios que celles qui avaient été étudiées dans le cadre du Baromètre des programmes radiophoniques de 2019[27]. En outre, comme en 2019, tous les programmes du corpus relèvent de la production propre des éditeurs, à l'exception de *Bruno sur Fun Radio* qui est produit par Fun Radio France.

Le corpus du volet « Focus » est composé de 176 heures de flux audio.

3.2. Unités d'encodage et de mesure

L'unité d'encodage du « Focus » est, comme pour le « Grand angle », la personne qui prend la parole au sein des programmes radiophoniques – alors que, comme expliqué précédemment, les Baromètres précédents s'intéressaient tant aux personnes qui parlent qu'à celles dont on parle (et, en télévision, celles que l'on voit). Pour rappel, les personnes prenant la parole représentaient 52,89% de l'ensemble des personnes encodées dans le Baromètre 2019.

Le volet « Focus » de la présente étude met en œuvre deux types de mesure distinctes pour quantifier la place des femmes, des hommes et des personnes non-binaires dans les émissions matinales analysées :

- d'une part, le nombre de personnes intervenant dans les émissions, soit le nombre de personnes différentes qui prennent la parole chaque jour dans les émissions de chaque chaîne du corpus[28] ;
- d'autre part, le temps de parole de chaque intervention parlée au sein des émissions analysées.

D'une part, quantifier le nombre de personnes qui prennent la parole permet de dresser des comparaisons avec les résultats du Baromètre 2019 – et ce, même si une différence de mesure subsiste. En effet, dans le Baromètre 2019,

[25] On a tiré au sort trois jours de la semaine et un jour du week-end afin de conserver, au sein de l'échantillon, les différences de programmation qui existent entre les jours de semaine et les jours de week-end. Cette réduction du corpus s'est avérée nécessaire pour assurer la faisabilité de l'étude au vu du temps requis pour l'encodage et la vérification manuelle du « Focus ».
[26] Comme c'était le cas dans le Baromètre 2019, on parle dans le « Focus », pour ces deux tranches horaires, de « matinales » ou d'« émissions matinales », étant donné que les pratiques d'écoute semblent décalées le week-end, où le pic d'audience se manifeste entre 10 et 14 heures. Il faut cependant noter que dans le « Grand angle », qui couvre toutes les tranches horaires, à la fois en semaine et le week-end, la tranche de 10 à 14 heures est nommée « milieu de journée » - que ce soit en semaine ou le week-end.
[27] Seule LN Radio ne figurait pas dans le corpus de 2019, la chaîne n'existant pas à l'époque. Par ailleurs, il faut relever que le corpus de 2019 comportait 7 jours consécutifs et non 4 jours d'une semaine reconstituée.
[28] Les personnes qui interviennent dans les émissions sont donc comptabilisées par chaîne et par jour alors que, dans les Baromètres précédents, elles étaient comptabilisées par émission. Cette différence est expliquée dans la note méthodologique de l'étude.

comme dans toutes les éditions précédentes, les personnes étaient répertoriées « une seule fois par émission, sans tenir compte de la fréquence ni de la durée de leur intervention » (CSA, 2019, p. 11). L'encodage de ce « Focus » se faisant sur la base non pas des émissions telles qu'elles sont décrites dans les grilles de programmes des radios, mais des séquences composant ces émissions[29], une comptabilisation par émission n'est pas possible ici. Une comptabilisation par séquence ne correspondant pas aux pratiques dominantes d'écoute de la radio, on a donc opté pour une comptabilisation par jour[30].

D'autre part, la mise en œuvre conjointe de ces deux unités de mesure permet une compréhension plus fine de la place des hommes et des femmes dans les émissions radiophoniques au travers de l'exploration de la relation entre ces deux unités de mesure.

3.3. Pré-annotation automatique par des modèles d'IA

Le corpus du volet « Focus » de cette étude a d'abord été traité par les mêmes modèles d'intelligence artificielle que le corpus du volet « Grand angle » :

- un modèle développé spécifiquement pour l'étude, permettant de distinguer les programmes, les morceaux musicaux et les spots publicitaires ;
- un modèle visant à catégoriser les voix entendues comme féminines ou masculines au travers de la création de vecteurs de voix uniques correspondant à chacune des voix contenues dans un flux audio ;
- un modèle permettant de transcrire automatiquement les paroles prononcées sous forme de texte[31].

Le corpus du « Focus » a en outre été soumis au traitement d'autres modèles d'IA.

Identification des tours de parole et des locuteurs et locutrices

L'un de ces modèles permet de diviser les flux audio en tours de parole distincts, en détectant quand une personne arrête de parler et quand quelqu'un d'autre prend la parole – c'est ce qu'on appelle la diarisation. Le flux audio est ainsi divisé en différents segments attribués chacun à un locuteur ou à une locutrice au travers de désignations génériques (premier locuteur, deuxième locuteur, etc.). Le modèle utilisé pour cette opération est le modèle de diarisation de la plateforme en libre accès Pyannote.

Après l'opération de diarisation, une deuxième opération vise à identifier le locuteur ou la locutrice qui intervient à chaque prise de parole. Pour ce faire, on utilise les vecteurs de voix créés par le modèle Ecapa (dans le cadre de la reconnaissance du genre), que l'on compare à une base de données de locuteurs et locutrices connu.es (ayant fait l'objet d'une identification lors de l'encodage manuel). Si le vecteur de voix d'un locuteur ou d'une locutrice du fichier en cours de traitement correspond à un vecteur de voix de la base de données, les interventions de cette personne sont automatiquement associées à l'identifiant de la base de données.

Cette étape a donc nécessité un travail d'encodage manuel préalable d'un ensemble de fichiers du corpus du « Focus » afin de nourrir la base de données des locuteurs et locutrices connu.es. Une évaluation des performances des modèles a ensuite été réalisée par Multitel conjointement pour les deux tâches – la diarisation et l'identification des locuteurs et locutrices. L'évaluation a été réalisée sur un sous-échantillon du corpus du « Focus », composé de 110 fichiers ayant été annotés manuellement pour la segmentation des émissions et les tours de parole des différent.es locuteurs et locutrices.

Dans le cadre de cette évaluation, la métrique qui a été calculée est l'exactitude, soit la proportion de prédictions correctes sur l'ensemble

[29] Et ce, afin de prendre en compte la forte hybridité des émissions radiophoniques par rapport aux émissions télévisuels, particulièrement marquée durant la tranche horaire matinale.
[30] Des analyses exploratoires ont en outre montré que ce mode de calcul ne modifiait que très faiblement les résultats de l'analyse par rapport à une comptabilisation par programme ni même par heure ou pour l'ensemble des 4 tranches horaires étudiées dans le corpus.
[31] Pour plus d'explications sur ces modèles, voir la partie consacrée au « Traitement automatique par des modèles d'IA » dans le cadre du « Grand angle ».

des prédictions effectuées[32]. Le taux d'exactitude global calculé pour la diarisation et l'identification des locuteurs et locutrices sur nos données est de 90,27%.

Si les performances apparaissent ici légèrement en-deçà des performances mesurées pour les autres modèles, elles n'en restent pas moins très bonnes pour ce type de tâche – d'autant plus que l'évaluation a été réalisée de la manière la plus « sévère » possible, en incluant notamment les passages où plusieurs voix se font entendre en même temps et où l'annotation automatique peut donc différer de l'annotation manuelle sans être pour autant fausse.

Contrairement au corpus du « Grand angle », le corpus du « Focus » a fait l'objet, après la phase de traitement automatique par les différents modèles d'IA, d'une vérification manuelle systématique – c'est pourquoi on parle ici de « pré-annotation automatique ».

3.4. Vérification et encodage manuel

L'équipe ayant réalisé ce travail a vérifié que tous les passages de programmes étaient bien identifiés comme tels. Au sein des programmes, elle a ensuite contrôlé l'identification des différents locuteurs et locutrices. Elle a également confronté la catégorisation des voix comme féminines ou masculines à ses propres perceptions du genre de la personne, en tenant compte de l'ensemble des indices disponibles (prénoms, utilisation des pronoms, accords grammaticaux, teneur des propos, etc.). A la suite de cette vérification, on parlera dans le volet « Focus » de l'étude d'hommes, de femmes, de personnes non-binaires et de personnes dont on n'a pas pu identifier le genre – tout en restant conscient qu'il s'agit bien de la perception des personnes ayant réalisé la vérification.

La vérification systématique des pré-annotations réalisées par les modèles d'IA a également été l'occasion d'enrichir le travail de catégorisation de l'IA au travers d'annotations manuelles qui nécessitent, à ce jour en tout cas, une analyse humaine. Il s'agit :

- d'une part, de segmenter les programmes du corpus en séquences et de classer ces différentes séquences en fonction du type de programme dont elles relèvent ;
- d'autre part, d'identifier le statut d'intervention des personnes qui prennent la parole dans les programmes et d'attribuer à chacune de ces personnes un rôle médiatique.

Type de programme et type de séquence

Comme le soulignait le Baromètre 2019, « les matinales radiophoniques sont des objets particulièrement hybrides dans leur construction » se composant « d'une multitude de séquences qui alternent et peuvent relever de genres ou de finalités énonciatives différentes » (CSA, 2019, p. 10). En 2024, cette hybridité est toujours une caractéristique essentielle des émissions constituant le corpus du « Focus ».

Les programmes du corpus ont dès lors été segmentés en séquences – soit des portions de programmes délimitées par un jingle, une coupure publicitaire, un titre musical ou une annonce explicite de l'équipe d'animation. Chaque séquence a ensuite été qualifiée comme relevant d'un type de programme, et au sein de celui-ci, d'un type de séquence, selon la typologie suivante[33] :

- **Animation générale :** toutes les séquences servant uniquement à faire le lien entre différents programmes, différentes séquences ou différents titres musicaux, principalement au travers d'annonces et de désannonces ;

[32] L'exactitude est une métrique plus pertinente que les métriques utilisées jusqu'à présent pour évaluer des modèles de classification avec de nombreuses catégories, comme c'est le cas ici – puisque chaque locuteur ou locutrice constitue une catégorie.

[33] En 2019, la méthodologie du Baromètre, conçue à l'origine pour la télévision, avait été adoptée à cette hybridité des programmes de la manière suivante : chaque programme analysé avait, dans son ensemble, été classé dans un genre (musical, information, magazine, divertissement) et un sous-genre médiatique (programme d'accompagnement et émission musicale, par exemple, pour le genre « musical »). Parallèlement à cette première classification, toutes les séquences composant les programmes analysés avaient été catégorisées selon leur format (annonce/désannonce, titres/flash info, info insolite…). Le volet « Focus » de la présente étude met en œuvre une approche différente ayant pour objectif de refléter au plus près le temps consacré aux différents genres médiatiques dans les matinales radiophoniques étudiées. La typologie utilisée dans ce volet de l'étude a été conçue comme une adaptation de la typologie des genres médiatiques mise en œuvre dans l'ensemble des Baromètres précédents tout en tenant compte des catégories de formats de séquences élaborées dans le Baromètre de 2019. On présente dans cette note l'ensemble des catégories et sous-catégories de cette typologie, même si certaines n'ont pas été mobilisées lors de l'encodage.

- **Musical**[34] **:**
    - Séquence sur la musique (peu importe le format : chronique, interview, portrait, agenda…) ;
    - Séquence d'accompagnement de la musique au travers d'interventions d'auditeurs ou d'auditrices ;
- **Information :**
    - Journal parlé et rappel des titres ;
    - Débat, forum ;
    - Entretien, interview ;
    - Autre séquence d'information (peu importe le format) ;
- **Magazine :**
    - Séquence sur la culture et le patrimoine (ne portant pas ou pas exclusivement sur la musique, peu importe le format de la séquence) ;
    - Documentaire ;
    - Séquence de société (peu importe le format) ;
    - Séquence consacrée au lifestyle (peu importe le format) ;
    - Autre (dont les séquences d'éducation permanente) ;
- **Divertissement et humour :**
    - Séquence humoristique (peu importe le format) ;
    - Retransmission d'un (extrait de) spectacle humoristique ;
    - Information insolite ;
    - Jeu ou jeu concours[35] ;
    - Interactions avec les auditeurs et auditrices sur un thème ne relevant ni de la musique ni de l'information ;
    - Talk-show ;
    - Autre ;
- **Sport :**
    - Séquence sur le sport (peu importe le format) ;
    - Retransmission d'un événement sportif ;
    - Autre ;
- **Info service :**
    - Météo ;
    - Info trafic ;
    - Autre ;
- **Autre.**

Des éléments d'habillage d'antenne peuvent être intégrés dans les émissions sans constituer pour autant des séquences[36]. Ces éléments ont été encodés de manière distincte en tant que :

- Jingles (comportant une ou plusieurs voix)[37] ;
- Bandes annonces et spots d'autopromotion.

Statut d'intervention

Les personnes prenant la parole dans les programmes analysés ont d'abord été qualifiées en fonction de leur statut par rapport à la chaîne de radio, soit le statut d'intervention.

Les différentes catégories de statut d'intervention utilisées dans le « Focus » sont :

- **Membre du personnel de la radio** (ou assimilé, comme les journalistes d'un autre média qui ont une chronique récurrente) ;
- **Personne invitée ou interviewée**[38] **;**
- **Membre du public intervenant directement**, c'est-à-dire les auditeurs et auditrices qui interagissent directement avec les personnes en charge de l'animation du programme ;

[34] Le type de programme « musical » a été utilisé pour toutes les séquences portant sur la musique (à l'exception des séquences se limitant à l'annonce ou la désannonce d'un titre musical, qui relèvent ici de l'« animation générale »). Les séquences portant sur des sujets culturels hors musique ont quant à elles été rattachées au « magazine ». Ce choix ne signifie bien évidemment pas que la musique n'est pas considérée comme faisant partie de la culture, mais entend simplement refléter l'importance de la musique en radio.

[35] En ce compris les séquences présentées sous la forme d'un jeu entre les membres de l'équipe d'animation, n'impliquant ni prix à gagner, ni auditeur ou auditrice pour participer.

[36] L'analyse ne portant que sur les programmes, seuls les jingles marquant le début ou la fin d'une séquence, ou inclus au sein d'une séquence, ont fait l'objet d'un encodage. De même, seuls les bandes annonces et spots d'autopromotions intégrés dans un programme ont été analysés. Par ailleurs, ces éléments d'habillage ne constituant pas des prises de parole, ils ont été traités de manière distincte lors de l'analyse (voir ci-dessous).

[37] La présente étude s'intéresser aux voix, les jingles n'en comportant pas n'ont pas été annotés.

[38] Afin d'éviter de « perturber » le modèle d'IA permettant la reconnaissance des voix des différents locuteurs et locutrices, une sous-catégorie a été créée, lors de l'encodage, pour les personnes interviewées ou invitées faisant l'objet d'un doublage par un ou une journaliste de la chaîne. Cette sous-catégorie a ensuite été intégrée à la catégorie des personnes interviewées ou invitées lors de l'analyse.

- **Membre du public intervenant indirectement**, soit les auditeurs et auditrices que l'on entend au travers de la diffusion de messages vocaux durant l'émission[39].

Certaines voix entendues dans les programmes ne correspondant à aucun de ces statuts, les passages où ces voix se font entendre ont été encodés en tant que[40] :

- **Fond sonore** (par exemple, une personne qui parle en arrière-fond) ou illustration sonore (par exemple, un extrait d'archives) ;
- **Extrait musical** (pour les extraits de titres musicaux comportant des voix et faisant partie d'une séquence de programme, comme lors d'annonces de titres à venir ou dans des chroniques musicales).

Rôle médiatique

Les personnes prenant la parole dans les programmes ont également été catégorisées en fonction du rôle qu'elles endossent au sein du dispositif que constitue l'émission – soit leur rôle médiatique. Il s'agit très largement des mêmes catégories que celles des Baromètres précédents[41] :

- **Journaliste, animateur ou animatrice :**
    - À titre principal dans le cadre de l'émission matinale ;
    - À titre principal dans le cadre des journaux parlés ou des rappels de titres ;
    - À titre secondaire (en charge d'un reportage, d'une chronique, d'une séquence particulière…) ;
- **Personne participant à un jeu ;**
- **Porte-parole**, en charge de représenter une personne, un groupe, une institution :
    - Porte-parole politique ;
    - Autre porte-parole ;
- **Personne intervenant pour son expertise,** soit une personne qui émet un avis sur la base d'une connaissance spécialisée ;
- **Personne intervenant à titre personnel :**
    - Quidam, qui émet un avis jugé comme étant le reflet du « citoyen ordinaire » ;
    - Témoin, qui s'exprime à titre personnel sur un sujet, émet un avis sur base de l'observation directe sans implication personnelle ;
    - Personne non connue faisant part d'une expérience personnelle, pour les personnes « ordinaires » dont l'intervention est centrée sur elles-mêmes, leur vécu, leurs réalisations personnelles (en ce compris les auditeurs et auditrices qui font une dédicace puisque celle-ci traduit leurs goûts personnels) ;
    - Personne connue faisant part d'une expérience personnelle, pour les interventions centrées sur le vécu et les réalisations personnelles des personnalités publiques ;
- **Autre.**

[39] Il faut souligner que nous utilisons donc les termes « auditeurs.trices direct.es » et « indirect.es » dans un sens différent de celui qu'ils revêtaient dans le Baromètre 2019. En effet, les « auditeurs.trices direct.es » y étaient défini.es comme les personnes issues du public prenant la parole à l'antenne tandis que les personnes du public qui ne prenaient pas la parole mais dont on parlait (par exemple, en lisant leurs messages postés sur les réseaux sociaux) y étaient considérées des « auditeurs.trices indirect.es » (CSA, 2019, p. 10)

[40] Ces passages ne renvoyant pas, à proprement parler, à des personnes prenant la parole dans les programmes, ils ont été traités de manière distincte dans l'analyse (voir ci-dessous).

[41] Par rapport au Baromètre des programmes radiophoniques de 2019, certaines catégories ont été supprimées, ayant perdu leur pertinence en raison des changements apportées à la méthodologie de cette nouvelle édition : « figurant.e », « voix d'habillage d'antenne » et « interprète de titre musical ». Une sous-catégorie a été ajoutée : « personne connue faisant part d'une expérience personnelle ». Cet ajout vise à déterminer la place accordée à l'antenne aux personnes partageant leur propre expérience selon qu'il s'agit de personnalités publiques ou non. Il a entraîné un changement de nom de la catégorie « vox populi » en « personne intervenant à titre personnel ».

3.5. Analyse de la présence et du temps de parole

Le cœur de l'analyse réalisée dans ce « Focus » porte sur les interventions parlées au sein des programmes composant la tranche horaire étudiée. Pour rappel, ce volet de l'étude met en œuvre deux unités de mesure distinctes pour quantifier la place des femmes, des hommes et des personnes non-binaires dans les émissions matinales analysées : d'une part, le nombre de personnes intervenant dans les émissions et d'autre part, le temps de parole de chaque intervention parlée.

Pour chacune de ces unités, les résultats sont analysés :

- en termes « absolus » afin d'appréhender les différences de volumes de données que l'on peut observer au sein du corpus en fonction des croisements effectués. On s'intéresse alors au nombre de femmes, d'hommes et de personnes non-binaires parmi les personnes qui interviennent et au temps de parole de chacun de ces groupes ;
- en termes « relatifs » afin de pouvoir comparer plus facilement la place des différents groupes. On s'intéresse alors au taux de présence, soit la part respective des femmes, des hommes et des personnes non-binaires au sein de l'ensemble des personnes qui interviennent dans les émissions, et au taux de parole, soit la part de temps de parole de chacun de ces groupes.
- Ces résultats sont présentés de manière globale, puis affinés :
- selon le type de programme et le type de séquence ;
- selon le statut de la personne qui prend la parole par rapport à la chaîne de radio ;
- selon le rôle médiatique que cette personne endosse au sein du programme.

Enfin, il faut noter que les jingles, les spots d'autopromotion, les extraits musicaux et les fonds et illustrations sonores faisant partie des émissions du corpus sont analysés de manière distincte. En effet, ces éléments ne constituant pas des prises de parole à proprement parler, ils ne font pas l'objet d'une analyse en termes de durée, mais uniquement en termes de nombre d'interventions[42] et de taux de présence selon que les voix qui y sont entendues sont des voix féminines, masculines ou mixtes (à la fois féminines et masculines).

[42] Ainsi, par exemple, des jingles comportant une voix identique ont été comptabilisés séparément à chacune de leur diffusion.

Annexes

Annexes

1. Liste des fichiers du corpus du « Grand angle » qui n'ont pas pu être traités en raison de problèmes techniques lors de l'enregistrement.

En 2024 :

- LN Radio : mercredi 11 septembre, à 2 heures
- Maximum FM : lundi 16 décembre, à 4 heures
- Vibration : dimanche 24 novembre, de 3 heures à 11 heures

En 2025 :

- Antipode : dimanche 9 novembre, à 22 heures et 23 heures
- BXFM : lundi 17 novembre, à 17 heures ; mercredi 3 décembre, à 7 heures ; jeudi 24 octobre, à 2 heures
- Maximum FM : lundi 17 novembre, à 23 heures ; mardi 30 septembre, à 8 heures, mercredi 3 décembre, à 8 heures et 10 heures
- NRJ : lundi 17 novembre, à 22 heures
- Classic 21 : lundi 17 novembre, à 22 heures
- VivaCité : lundi 17 novembre, à 22 heures
- Radio Contact : lundi 17 novembre, à 22 heures
- Judaïca : mercredi 3 décembre, à 21 heures et 22 heures

2. Liste des fichiers du corpus du « Grand angle » qui n'ont pas pu être traités dans leur entièreté en raison de problèmes techniques lors de l'enregistrement.

En 2024 :

- Antipode : vendredi 30 août, à 11 heures
- Equinoxe FM : lundi 16 décembre, à 10 heures
- LN Radio : mercredi 11 septembre, à 18 heures
- Maximum FM : jeudi 24 octobre, à 8 heures
- Inside Radio : vendredi 30 août, à 13 heures
- Nostalgie : lundi 16 décembre, à 1 heure ; vendredi 30 août, à 9 heures
- RCF Liège : vendredi 30 août, à 9 heures
- Musiq3 : lundi 16 décembre, à 6 heures
- Sud Radio : vendredi 30 août, à 9 heures

En 2025 :

- Antipode : mardi 30 septembre, à 22 heures ; dimanche 9 novembre, à 21 heures
- Fun Radio : jeudi 4 septembre, à 2 heures
- Vibration : lundi 17 novembre, à 2 heures

BAROMÈTRE

CSA
CONSEIL SUPÉRIEUR DE L'AUDIOVISUEL

septembre 2026