

## **AUDIMETRIE PORTEE ET PRISE EN COMPTE DES NOUVEAUX MODES DE CONSOMMATION RADIO : EXEMPLE DE L'ECOUTE AU CASQUE**

Lorie DUDOIGNON (\*), Flavien ALLEAUME (\*\*)

(\*) Médiamétrie, Directrice de l'innovation scientifique

(\*\*) Médiamétrie et Insee – Direction Régionale de Normandie

lorie.dudoignon@mediametrie.fr

**Mots-clés:** Audience Radio, Ecoute au casque, Fusion de données, Imputation

**Domaine concerné :** Appariements, fusion de sources (record linkage), couplage « exact » de fichiers

---

### **Résumé**

*La mesure de référence d'audience Radio en France est réalisée par Médiamétrie. Elle s'appuie sur deux dispositifs complémentaires :*

*Le premier appelé EAR > National – Etude Audience Radio National est une enquête réalisée par téléphone auprès de 100.000 individus de 13 ans et plus de septembre à juin. Dans cette enquête déclarative, l'audience Radio des dernières 24 heures est recueillie par demi quart d'heure.*

*Le second dispositif doit permettre de suivre les comportements d'écoute Radio d'un même individu dans le temps à la fois pour des analyses éditoriales mais aussi et surtout pour alimenter les outils de médiaplanning Radio (c'est-à-dire pour permettre aux agences, annonceurs de définir leur plan média pour la diffusion d'une campagne publicitaire).*

*Jusqu'à aujourd'hui ce dispositif appelé Panel Radio était une enquête réalisée via des carnets d'écoute en auto-administré. En septembre 2022, ce dispositif sera remplacé par l'étude EAR > Insights s'appuyant sur un panel (appelé AIP pour Audimétrie Individuelle Portée) permanent de 5 000 individus âgés de 13 ans et plus, équipés d'un audimètre miniature les accompagnant dans leurs activités quotidiennes. Basé sur la technologie du watermarking (marque inaudible insérée par les éditeurs, dans le son des radios et TV), cet audimètre, porté à la ceinture ou au poignet par les panélistes, identifie les écoutes réalisées par les panélistes tout au long de la journée.*

*Ce nouveau dispositif présente de nombreux avantages par rapport à l'ancien dispositif du Panel Radio. La mesure étant permanente elle permettra de mieux refléter la saison radiophonique, avec des résultats disponibles plus fréquemment. Une mesure automatique offre aussi des possibilités de mesure plus fine et plus précise des comportements d'écoute Radio que ce que la mémoire des individus permet de restituer avec un carnet d'écoute.*

*Les écoutes réalisées via un casque sont théoriquement mesurables grâce à un adaptateur fourni aux panélistes. Néanmoins, cet adaptateur est très peu utilisé par les panélistes car son utilisation est jugée contraignante, les écoutes au casque ne sont que très partiellement mesurées.*

*Nous sommes donc en présence de données manquantes que nous souhaitons imputer.*

*L'approche proposée pour cette imputation est une fusion de type donneur-receveur dans laquelle l'enquête EAR > National servira de base donneuse grâce à l'ajout d'un module au questionnaire permettant d'identifier les écoutes au casque.*

*Le principe est le suivant :*

*Pour chaque individu i de la base receveuse, on cherche un donneur dans la base donneuse. Les données observées pour ce donneur vont être transférées à son receveur. Chaque receveur est unique, mais un donneur peut potentiellement être attribué plusieurs fois. Pour déterminer si deux individus se ressemblent, toutes les variables communes peuvent être utilisées, soit pour calculer une distance, soit pour définir des strates.*

*Dans notre cas :*

- *La base receveuse correspond aux individus du panel AIP*
- *La base donneuse correspond aux individus de l'enquête EAR > National*
- *Les données communes sont les variables socio-démographiques et les audiences Radio hors casque*
- *Les données à transférer sont les audiences Radio au casque*

## **Abstract**

*Médiamétrie has set up an automatic measurement panel for radio audience: the AIP panel. Panelists are asked to carry a portable meter with them, anytime and anywhere. This meter can recognize the TV channels watched and the Radio stations listened by the panelists, whatever the location, thanks to the watermarking audio technology. The Watermarking consists in inserting an inaudible "code" in the audio signal of TV channels and Radio stations. This code can be recognized by our meters.*

*For listening with headphones, an adapter is provided to the panelists, but its complexity of use means that in practice, few use it. So we have some missing data in this specific case. We propose a solution of data fusion to impute radio listening with headphones.*

*The second source of data we will use in this data fusion is a declarative survey in which we collect all radio audience for the last 24 hours with the detail of listening with headphones or not.*

---

## **1. Introduction**

La mesure de référence d'audience Radio en France est réalisée par Médiamétrie. Elle s'appuie sur deux dispositifs complémentaires :

- Le premier appelé EAR > National – Etude Audience Radio National (anciennement 126 000 Radio) est une enquête réalisée par téléphone auprès de 100.000 individus de 13 ans et plus de septembre à juin. Dans cette enquête déclarative, l'audience Radio des dernières 24 heures est recueillie par demi quart d'heure.
- Le second dispositif doit permettre de suivre les comportements d'écoute Radio d'un même individu dans le temps à la fois pour des analyses éditoriales mais aussi et surtout pour alimenter les outils de médiaplanning Radio (c'est-à-dire pour permettre aux agences, annonceurs de définir leur plan média pour la diffusion d'une campagne publicitaire).

Jusqu'à aujourd'hui ce dispositif appelé Panel Radio était une enquête réalisée auprès de 10.000 individus 13 ans et plus en auto-administré. Les individus recrutés pour le Panel Radio devaient chacun remplir un carnet d'écoute (papier ou digital) durant 23 jours consécutifs. En septembre 2022, ce dispositif sera remplacé par l'étude EAR > Insights s'appuyant sur un panel (appelé AIP pour Audimétrie Individuelle Portée) permanent de 5 000 individus âgés de 13 ans et plus, représentatifs de la population française, équipés d'un audimètre miniature les accompagnant dans leurs activités quotidiennes. Basé sur la technologie du watermarking (marque inaudible insérée par les éditeurs, dans le son des radios et TV), cet audimètre, porté à la ceinture ou au poignet par les panélistes, identifie les écoutes réalisées par les panélistes tout au long de la journée. Ce nouveau dispositif présente de nombreux avantages par rapport à l'ancien dispositif du Panel Radio. La mesure étant permanente elle permettra de mieux refléter la saison radiophonique, avec des résultats disponibles plus fréquemment (jusqu'à 4 vagues par an vs 1 actuellement avec le Panel Radio), une plus grande

profondeur d'analyse (des périodes de 4 semaines vs 23 jours actuellement) et la possibilité de disposer d'analyses éditoriales Ad Hoc tout au long de l'année. Une mesure automatique offre aussi des possibilités de mesure plus fine et plus précise des comportements d'écoute Radio que ce que la mémoire des individus permet de restituer avec un carnet d'écoute.

Les écoutes réalisées via un casque sont théoriquement mesurables grâce à un adaptateur fourni aux panélistes. Néanmoins, cet adaptateur est très peu utilisé par les panélistes car son utilisation est jugée contraignante, les écoutes au casque ne sont que très partiellement mesurées.

Une petite partie des écoutes radio sont donc manquantes (environ 5% des écoutes radio sont réalisées via casque et cet usage concerne un peu plus de 5% de la population âgée de 13 ans et plus au quotidien, en semaine) et nous devons les imputer.

## 2. Généralités sur la fusion

L'approche proposée est une fusion de type donneur-receveur. Si on s'oriente vers cette solution, c'est parce qu'avec cette approche la base receveuse est complétée mais pas modifiée ce qui garantit une conservation des résultats du panel AIP sur la partie Audience Hors Casque – soit la majeure partie des usages.

La solution repose sur la complémentarité des deux dispositifs de mesure radio décrits en introduction. Un module spécifique permettant de recueillir en détail l'écoute au casque a été ajouté à l'enquête EAR > National sur les périodes Septembre-Octobre et Janvier-Mars, pour servir de base donneuse à la fusion.

### 2.1. Des temporalités différentes

Selon les productions EAR > Insights nous devons réaliser l'imputation pour 3 ou 4 semaines consécutives. Dans la suite, on considèrera un bloc de 4 semaines à titre d'exemple mais le principe est identique pour un bloc de 3 semaines. L'imputation doit permettre d'ajouter de l'audience au casque sur l'ensemble de la période étudiée. Or dans l'enquête donneuse, nous ne disposons de données que pour une journée par individu. Pour que l'imputation soit complète nous aurons donc besoin pour chaque individu receveur de la base AIP de plusieurs donneurs issus de l'enquête déclarative.

Nous avons choisi d'associer à chaque individu receveur 7 donneurs : 1 par jour nommé<sup>1</sup>. Ce qui nous permet tout en tenant compte des spécificités de chaque jour de la semaine de :

- Limiter le nombre de donneurs différents pour un même receveur afin de ne pas multiplier « artificiellement » le nombre de stations écoutées par un individu sur la période
- Utiliser l'ensemble des interviews d'un jour nommé comme base donneuse et pas seulement celles d'une date ce qui réduirait considérablement le nombre de donneurs potentiels et donc la qualité de la fusion

Pour chaque individu  $i$  x jour nommé de la base receveuse, on cherche donc un donneur dans la base donneuse (pour ce même jour nommé). Les données observées pour ce donneur vont être transférées à son receveur pour les 4 jours considérés (si on considère un bloc de 4 semaines).

### 2.2. Grandes étapes

Pour que la fusion ait du sens, il faut que les couples (receveur, donneur) formés correspondent à des individus qui se « ressemblent ». Pour déterminer si deux individus se ressemblent, toutes les variables communes peuvent être utilisées, soit pour calculer une distance, soit pour définir des strates. La fusion s'effectue alors en 4 opérations successives : sélection des individus éligibles, stratification, calcul de la distance et jumelage.

Chaque receveur est unique, mais un donneur peut potentiellement être attribué plusieurs fois. Le nombre de réplification d'un même donneur est un critère à contrôler lors de l'étape du jumelage.

---

<sup>1</sup> Lundi, Mardi ..., Dimanche

### 3. Individus éligibles et stratification

#### 3.1. Individus éligibles à la fusion

Nous disposons pour cette imputation des audiences Radio au casque d'une information précieuse de fréquence d'écoute au casque. Cette question est posée à la fois aux panélistes AIP et aux interviewés EAR > National. Les périodes pour lesquelles nous souhaitons faire une imputation étant de 4 semaines maximum, seuls les individus déclarant avoir des audiences Radio via le casque plus d'une fois par mois seront éligibles à la fusion (que ce soit du côté des donneurs et des receveurs). Les individus AIP ayant des audiences casque mesurées par l'audimètre et donc utilisant correctement l'adaptateur fourni à cet effet sont eux aussi écartés de la fusion puisqu'ils n'ont pas besoin d'imputation.

#### 3.2. Stratification

L'objectif principal de la stratification est d'interdire certains jumelages. La fusion étant ensuite appliquée par strate de manière indépendante, cela permet aussi d'optimiser les temps de calcul (parallélisation possible des calculs).

Nous n'avons pas des effectifs suffisants côté receveur pour mettre en place un nombre important de strates. En effet, il est certes séduisant d'avoir des receveurs et des donneurs très similaires d'un point de vue socio-démographique, mais cela se ferait forcément au détriment de la variété des comportements répliqués, et de la proximité (distance) constatée entre les donneurs et les receveurs. Mieux vaut donc avoir des strates un peu plus grosses, mais avec des comportements d'écoute plus ressemblants.

Nous avons fait le choix de croiser pour les strates les 2 variables qui sont les plus explicatives du comportement au casque, à savoir :

- Les habitudes d'écoute au casque déclarées - trivial, mais central !
- L'âge - qui non seulement est très corrélé avec l'usage du casque mais aussi un critère très discriminant sur les usages Radio dans leur ensemble comme les stations écoutées notamment : plus musicales chez les jeunes par exemple ...

Les strates obtenues sont au nombre de 8 et sont les suivantes :

- 13-24 ans assidus au casque
- 25-34 ans assidus au casque
- 35-49 ans assidus au casque
- 50 ans et plus assidus au casque
- 13-24 ans occasionnels du casque
- 25-34 ans occasionnels du casque
- 35-49 ans occasionnels du casque
- Plus de 50 ans occasionnels du casque

#### 3.3. Retrait des « atypiques »

Lors de la phase de R&D, nous avons pu constater que l'effet d'un individu atypique parmi les donneurs peut être très important sur les résultats finaux, et ce d'autant plus que celui-ci serait affecté à plusieurs receveurs. La stabilité des résultats étant primordiale, nous avons introduit la possibilité de filtrer ces individus atypiques.

Un individu sera considéré comme atypique s'il fait partie des x% les plus gros consommateurs au casque du jour. Le x% est défini par strate (seuil paramétrable) et peut être différent selon les strates.

En première approche, nous avons choisi de retirer 1% des individus les plus consommateurs de casque sur les 4 strates d'assidus.

## 4. Fonction de distance

### 4.1. Création des variables pour la distance

Les variables retenues pour la fonction de distance correspondent aux audiences hors casque (sous forme de contact oui/non) des 23 stations (ou agrégats de station) utilisées dans le Médiaplanning Radio croisées par 5 grandes tranches horaires (0h-6h/6h-9h/9h-14h/14h-18h/18h-0h). On obtient ainsi un total de 115 variables 0/1 permettant de calculer les distances entre donneurs et receveurs. Ces variables ont été choisies selon un critère « métier » afin d'assurer la cohérence des audiences et des profils d'auditoire.

### 4.2. Choix de la fonction de distance

Initialement, nous nous étions orientés vers la distance de Jaccard :

$$J = 1 - \frac{n_{11}}{n_{11} + n_{10} + n_{01}}$$

où  $n_{11}$  = nombre de stations × TH validés dans EAR – National et dans l'AIP

$n_{10}$  = nombre de stations × TH validés dans EAR – National mais pas dans l'AIP

$n_{01}$  = nombre de stations × TH validés dans l'AIP mais pas dans EAR – National

Cependant, nous constatons quelques différences entre les niveaux d'audience hors casque mesurés dans EAR > National et dans l'AIP. Ces différences s'expliquent par le fait que les modes de recueil sont de nature très différente avec un déclaratif au demi-quart d'heure dans EAR > National versus une mesure automatique dans le panel AIP.

Nous avons donc choisi de modifier la distance de Jaccard de la manière suivante :

$$J = 1 - \frac{n_{11} + \alpha n_{10}}{n_{11} + n_{10} + n_{01}}$$

On introduit la pondération  $\alpha \in [0,1]$  afin de l'imiter l'impact de  $n_{10}$  par rapport à  $n_{11}$ .

$\alpha = 0$  : distance de Jaccard classique

$\alpha = 1$  :  $n_{11}$  et  $n_{10}$  ont la même importance

Avec  $\alpha = 0$ , les non-auditeurs « hors casque » EAR-National ont tendance à être plus utilisés que les autres donneurs (nombre de réplifications plus important que la moyenne) du fait qu'ils sont proportionnellement moins nombreux dans l'enquête déclarative que dans la mesure automatique.

Avec  $\alpha = 0.5$ , le nombre de réplifications des non-auditeurs « hors casque » EAR > National est en moyenne identique aux autres donneurs. De plus, le taux d'audience simultanée casque et hors-casque après fusion est diminué.

Nous avons donc retenu cette valeur pour  $\alpha$ .

## 5. Algorithme de jumelage

Le jumelage doit remplir les conditions suivantes :

- Associer des individus qui se ressemblent (en minimisant la distance définie en 4.)
- Associer à chaque receveur un unique donneur
- Utiliser tous les donneurs au moins une fois (sauf s'il y a plus de donneurs que de receveurs, alors certains donneurs ne seront pas utilisés)
- Si certains donneurs sont attribués plusieurs fois, limiter leur nombre d'attributions : il vaut mieux utiliser des donneurs différents plutôt que d'utiliser plusieurs fois le même donneur (même s'il est plus proche des receveurs) – Si on utilise tous les donneurs un nombre identique de fois on conservera les résultats obtenus sur la base donneuse. A contrario, plus la distribution du nombre de réplification est dispersée et plus on s'éloigne des résultats de la base donneuse.

Pour répondre à cet objectif, l'algorithme Least Cost Method (résolution d'un problème de transport) a été sélectionné et légèrement adapté aux besoins cités ci-dessus.

### 5.1. Notations

$i$  : un donneur (panéliste EAR > National)

$n_D$  : nombre de donneurs

$j$  : un receveur (panéliste AIP)

$n_R$  : nombre de receveurs

$d_{i,j}$  : distance entre le donneur  $i$  et le receveur  $j$

$QD_i$  : Vecteur contenant les quantités à donner, c'est-à-dire le nombre de fois que l'on peut attribuer le donneur  $i$

$QR_j$  : Vecteur contenant les quantités à recevoir, c'est-à-dire le nombre de donneurs à associer à un receveur  $j$

### 5.2. Etapes de l'algorithme de jumelage

Le fonctionnement de l'algorithme est décrit par étapes qui sont détaillées ci-dessous.

A noter que l'algorithme peut être réalisé en une ou deux phases :

Si la strate contient au moins autant de donneurs que de receveurs, il suffira d'une seule phase pour réaliser le jumelage.

Si la strate contient moins de donneurs que de receveurs, on utilisera tous les donneurs une fois, puis on réalisera une 2e phase pour attribuer des donneurs aux receveurs pour lesquels le jumelage n'a pas été réalisé à l'issue de la première phase.

#### 5.2.1. Initialisation phase 1

Techniquement, l'algorithme Least Cost Method (LCM) peut être amené à faire des choix en se basant sur l'ordre de la matrice. Afin de remplir le critère de reproductibilité de l'algorithme, on effectue une première étape d'initialisation qui consiste à trier la matrice de distance en fonction des identifiants des donneurs  $i$  et ceux des receveurs  $j$ .

Matrice de distance - Initialisation

Donneurs	Receveurs			
	$j_1$	$j_2$	...	$j_{n_R}$
$i_1$	$d_{1,1}$	$d_{1,2}$	...	$d_{1,n_R}$
$i_2$	$d_{2,1}$	$d_{2,2}$	...	$d_{2,n_R}$
...	...	...	...	...
$i_{n_D}$	$d_{n_D,1}$	$d_{n_D,2}$	...	$d_{n_D,n_R}$

On initialise la matrice d'allocation correspondant à une matrice de taille identique à la matrice de distance. Cette matrice permettra de stocker quel donneur est affecté à quel receveur.

Matrice d'allocation - Initialisation

Donneurs	Receveurs			
	$j_1$	$j_2$	...	$j_{n_R}$
$i_1$	$x_{1,1}$	$x_{1,2}$	...	$x_{1,n_R}$
$i_2$	$x_{2,1}$	$x_{2,2}$	...	$x_{2,n_R}$
...	...	...	...	...
$i_{n_D}$	$x_{n_D,1}$	$x_{n_D,2}$	...	$x_{n_D,n_R}$

A l'initialisation, tous les  $x_{i,j}$  sont égaux à 0.

A la fin de l'exécution du jumelage, on aura :

$$x_{i,j} = \begin{cases} 1 & \text{si le donneur } i \text{ a été attribué au receveur } j \\ 0 & \text{sinon} \end{cases}$$

La quantité à recevoir  $QR_j$  correspond au nombre de fois où chaque receveur doit être associé à un donneur. Ici, on veut seulement un donneur par receveur donc  $QR_j$  sera toujours égal à 1.

La quantité à recevoir  $QD_i$  correspond au nombre de fois où chaque donneur doit être attribué. Pour utiliser le maximum de donneurs possibles au moins une fois lors de la première phase, on initialise les  $QD_i$  à 1.

Matrice de distance - Ajout des quantités à donner et à recevoir

Donneurs	Receveurs				Quantité à donner ( $QD_i$ )
	$j_1$	$j_2$	...	$j_{n_R}$	
$i_1$	$d_{1,1}$	$d_{1,2}$	...	$d_{1,n_R}$	$QD_1 = 1$
$i_2$	$d_{2,1}$	$d_{i_2,j_2}$	...	$d_{2,n_R}$	$QD_2 = 1$
...	...	...	...	...	...
$i_{n_D}$	$d_{n_D,1}$	$d_{n_D,2}$	...	$d_{n_D,n_R}$	$QD_{n_D} = 1$
Quantité à recevoir ( $QR_j$ )	$QR_1 = 1$	$QR_2 = 1$	...	$QR_{n_R} = 1$	

L'algorithme est ensuite itératif. Le principe est de trouver un premier couple (receveur, donneur), puis pour un 2e, etc. Chaque itération est décrite dans les étapes ci-dessous.

### 5.2.2. Attribution d'un donneur à un receveur phase 1

Dans la matrice de distance, trouver la cellule  $(i, j)$  pour laquelle la distance  $d_{i,j}$  est la plus petite.

- En cas d'égalité sur la valeur minimale, trouver le donneur  $i$  avec la quantité  $QD_i$  la plus élevée. Lors de la première itération,  $QD_i$  est toujours égal à 1 : passer à la condition suivante.
- En cas d'égalité sur la quantité  $QD_i$  maximale, prendre le premier donneur  $i$  (tri par l'identifiant) qui respecte les conditions précédentes.

Dans la matrice d'allocation, allouer à la cellule  $(i, j)$  la valeur suivante :  $x_{i,j} = \min(QD_i, QR_j)$

### 5.2.3. Modification de la matrice de distance phase 1

A l'issue de l'étape précédente, un donneur a été attribué à un receveur. On modifie alors la matrice de distance de la manière suivante :

1. On modifie les quantités à donner et à recevoir :
  - $QD_i = QD_i - 1$
  - $QR_j = QR_j - 1 = 0$
2. On retire de la matrice de distance la colonne correspondant au receveur  $j$  à qui on a attribué un donneur.
3. Si  $QD_i = 0$  (ce qui est le cas lors de la 1ère phase), on retire également la ligne correspondant au donneur  $i$  qui a été attribué (et qui ne pourra plus être attribué aux receveurs restants lors de la phase en cours).

On itère ensuite les étapes 5.2.2 et 5.2.3 jusqu'à ce que :

- $\sum_{j=1}^{n_R} QR_j = 0$  : tous les receveurs ont un donneur (la phase 2 sera alors inutile)
- $\sum_{i=1}^{n_D} QD_i = 0$  : tous les donneurs ont été attribués une fois mais il reste des receveurs sans donneur. On doit passer à la phase 2.

### 5.2.4. Phase 2

Lors de la seconde phase, la matrice de distance doit se restreindre aux receveurs qui n'ont bénéficié d'aucune allocation lors de la première phase. On repart de la matrice de distance de départ et on la réduit à une matrice de taille  $(n_D, n_R - n_D)$ .

Le nouveau vecteur des quantités à recevoir  $QR$  reste composé uniquement de 1.

Le nouveau vecteur des quantités à donner  $QD$  prend toujours la même valeur, égale à l'arrondi à l'entier inférieur de  $(n^R/n_D + p)$  où  $p$  est un entier supérieur ou égal à 0 (par défaut à 0). Cela doit permettre d'assurer le fait que le nombre de donneurs disponibles est supérieur au nombre de receveurs restants, cela afin d'être sûr que chaque receveur aura un donneur attribué à la fin de la phase 2.

Exemple : il reste les receveurs 1, 3, 7 et 10 qui n'ont pas encore de donneur. Les matrices de distance et d'allocation seront donc de taille  $(n_D, 4)$ .

Matrice de distance – Initialisation pour la 2<sup>ème</sup> phase

Donneurs	Receveurs				Quantité à donner ( $QD_i$ )
	$j_1$	$j_3$	$j_7$	$j_{10}$	
$i_1$	$d_{1,1}$	$d_{1,3}$	$d_{1,7}$	$d_{1,10}$	$QD_1 = \text{ent}(n^R/n_D + p)$
$i_2$	$d_{2,1}$	$d_{2,3}$	$d_{2,7}$	$d_{2,10}$	$QD_2 = \text{ent}(n^R/n_D + p)$
...	...	...	...	...	...
$i_{n_D}$	$d_{n_D,1}$	$d_{n_D,3}$	$d_{n_D,7}$	$d_{n_D,10}$	$QD_{n_D} = \text{ent}(n^R/n_D + p)$
Quantité à recevoir ( $QR_j$ )	$QR_1 = 1$	$QR_3 = 1$	$QR_7 = 1$	$QR_{10} = 1$	

Matrice d'allocation – Initialisation pour la 2<sup>ème</sup> phase

Donneurs	Receveurs			
	$j_1$	$j_3$	$j_7$	$j_{10}$
$i_1$	$x_{1,1}$	$x_{1,3}$	$x_{1,7}$	$x_{1,10}$
$i_2$	$x_{2,1}$	$x_{2,3}$	$x_{2,7}$	$x_{2,10}$
...	...	...	...	...
$i_{n_D}$	$x_{n_D,1}$	$x_{n_D,3}$	$x_{n_D,7}$	$x_{n_D,10}$

avec  $x_{i,j} = 0$ .

Il suffit ensuite de refaire tourner les étapes 5.2.2 et 5.2.3 avec les nouvelles matrices de distance et d'allocation jusqu'à ce que tous les receveurs aient un donneur attribué.

### 5.2.5. En sortie de l'algorithme

En sortie, créer la table des couples sélectionnés (en réunissant les deux matrices d'allocation s'il y a eu 2 phases). Pour chaque receveur  $j$ , attribuer le donneur dont l'allocation  $x_{i,j} = 1$ .

Matrice d'allocation finale

Donneurs	Receveurs			
	$j_1$	$j_2$	...	$j_{n_R}$
$i_1$	0	1	...	0
$i_2$	0	0	...	1
...	...	...	...	...
$i_{n_D}$	0	0	...	0

Résultat du jumelage

Receveurs	Donneurs	Distance du couple retenu
$j_1$	$i_3$	$d_{3,1}$
$j_2$	$i_1$	$d_{1,2}$
...	...	...
$j_{n_R}$	$i_2$	$d_{2,n_R}$



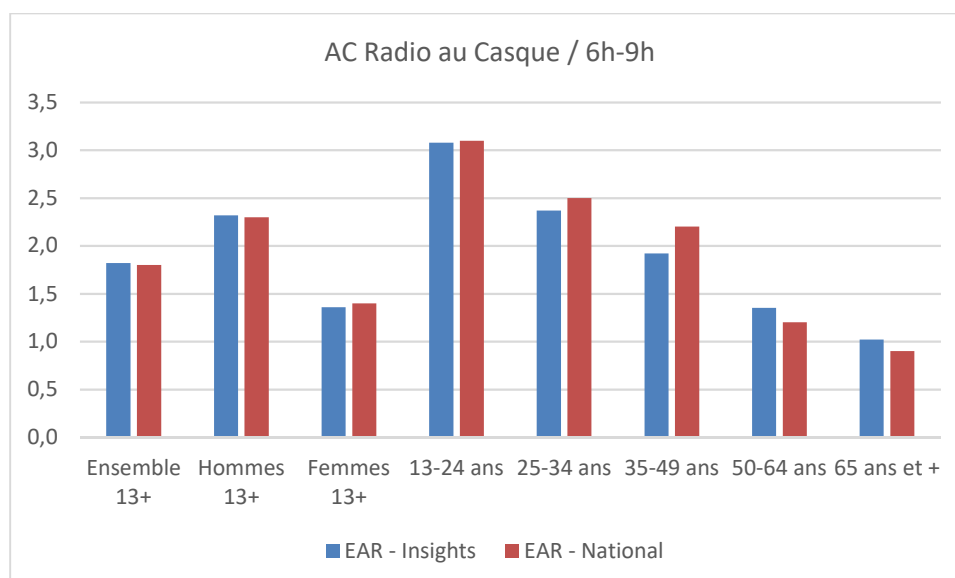
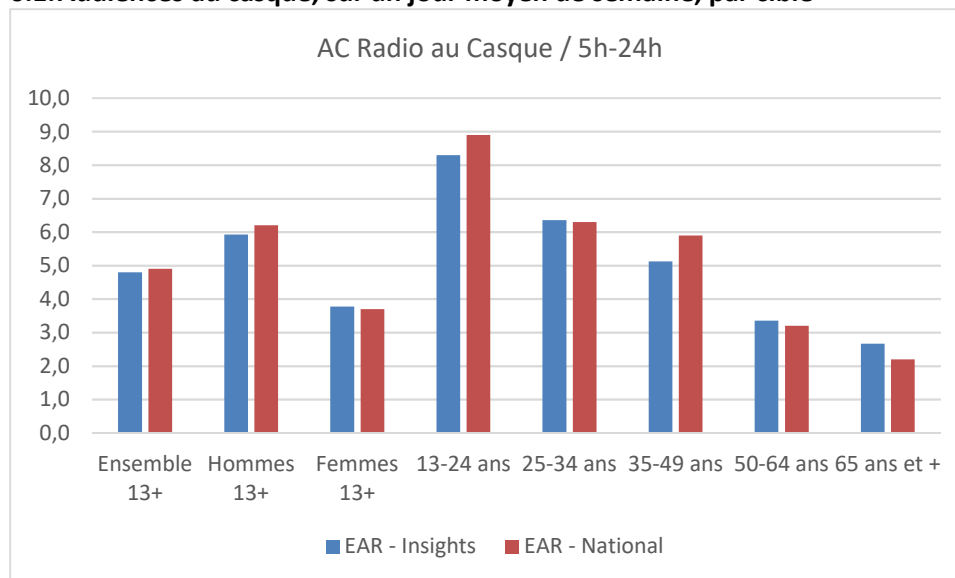
## 6. Résultats

La méthode décrite dans cet article est très proche de celle utilisée pour l'ajout des audiences hors-domicile en TV [2] et sur laquelle nous avons à présent près de 2 ans de recul. Nous savons que les algorithmes utilisés sont très robustes et donnent des résultats tout à fait acceptables en termes d'optimisation des distances et de contrôle des réplifications. Bien sûr, l'approche doit être adaptée au contexte tout à fait différent de la Radio et en particulier sur la fonction de distance.

La première publication de résultats EAR >Insights est prévue pour l'automne 2022 sur la période Septembre-Octobre 2022. Nous sommes actuellement en phase de test et de petits ajustements sont encore possibles. Néanmoins, nous pouvons partager quelques résultats des derniers tests en date.

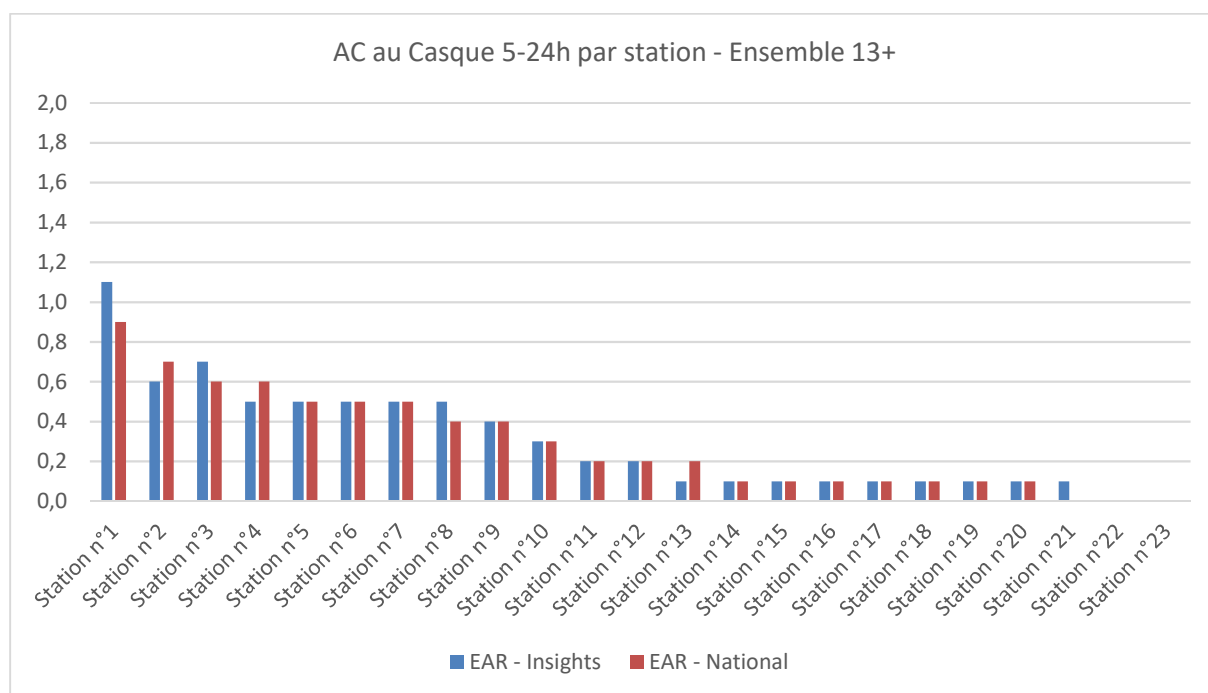
L'indicateur d'audience le plus utilisé en Radio est l'AC (pour Audience Cumulée) et correspond aux % d'auditeurs. C'est donc principalement sur cet indicateur que nous analysons les résultats obtenus. Nous comparons les audiences au casque mesurées dans l'enquête déclarative à celles obtenues dans EAR >Insights à l'issue de la fusion. La période analysée est Septembre-Octobre 2021.

### 6.1. Audiences au casque, sur un jour moyen de semaine, par cible



On observe que les différences d'usage liées au genre ou à l'âge sont conservées par la fusion – que ce soit globalement sur l'ensemble de la journée ou sur la tranche matinale (correspondant au « prime » du média radio).

## 6.2. Audiences au Casque, sur un jour moyen de semaine, par station



Les niveaux d'audiences au casque observés dans EAR > National sont globalement conservés et les écarts constatés sont de l'ordre de l'arrondi – les résultats étant publiés avec une décimale.

Si nous analysons ces usages en détail c'est pour s'assurer que le mécanisme d'imputation ne conduirait pas à une sur ou sous-estimation de certains usages spécifiques. Néanmoins, il faut garder en tête que ces résultats n'ont pas vocation à être publiés. Seule l'audience globale (avec et sans casque) sera disponible pour le marché et compte-tenu de la part relativement faible des écoutes au casques (environ 5% des écoutes Radio), l'effet de la fusion sur les résultats publiés restera tout à fait marginal. Les résultats actuels sont dans ce contexte tout à fait satisfaisants.

## Bibliographie

- [1] Dudoignon, L. (2018). « Fusion statistique de données d'enquête : dernières avancées pour les mesures d'audience », 10ème Colloque Francophone sur les Sondages, Lyon.
- [2] Dudoignon, L. (2021). « Fusion statistique de données d'enquête : cas des audiences TV Hors-Domicile », 11ème Colloque Francophone sur les Sondages, Bruxelles.
- [3] Fischer, N. (2004), « Fusion statistique de fichiers de données », Thèse de doctorat, Montpellier.
- [4] Mansi, S.G. (2011), « A Study on Transportation Problem, Transshipment Problem, Assignment Problem and Supply Chain », Thèse de doctorat, Rajkot.
- [5] Tassi, P. (2005), « Modèles statistiques de la mesure d'audience des médias audiovisuels », *Economica*.