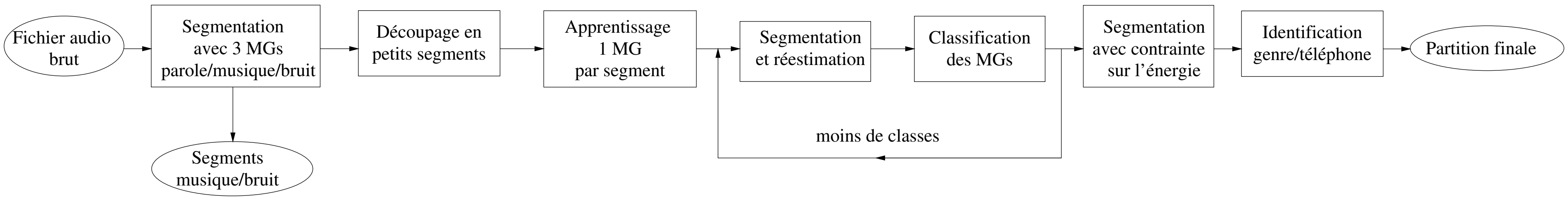

FIG. 1 – *Partition du signal audio. La segmentation consiste à diviser les données en segments acoustiques homogènes, c'est-à-dire à placer les frontières des segments là où il y a un changement significatif dans les conditions acoustiques (musique, bruit, parole téléphonique ou non) ou lorsqu'il y a un changement de locuteur (tour de parole). Pendant l'étiquetage, on associe une étiquette à chaque segment de l'émission, cette étiquette tenant compte du locuteur, du genre et de certaines conditions acoustiques*

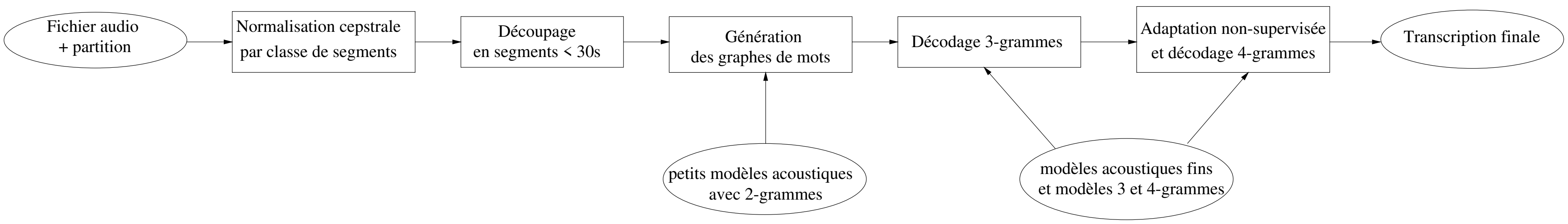

FIG. 2 – *Décodage par mot. La transcription orthographique est obtenue en trois passes : 1) génération d'un graphe de mots, 2) décodage avec un modèle 3-gramme, 3) décodage avec les modèles acoustiques adaptés. La sortie du système de reconnaissance consiste en la suite de mots, avec pour chaque mot des balises temporelles de début et de fin et un score de confiance.*