# Amélioration de la Reconnaissance Vocale pour la Documentation Clinique : Création d'un Corpus Audio Représentatif des Accents Ouest-Africains et Adaptation de Modèles d'IA

By Houessouvo C. Roland¹,2, Keke Mahuvivi Turibio3, Jossou R. Thierry¹,2, Medenou Daton¹,2

*¹ Department of Biomedical Engineering (GBM), École Polytechnique d’Abomey-Calavi (EPAC), Université d’Abomey-Calavi, Abomey-Calavi, 01 BP 2009 Cotonou, Benin  
 ² Laboratoire d’Electrotechnique, de Télécommunication et d’Informatique Appliquée (LETIA), Université d’Abomey-Calavi, Abomey-Calavi, 01 BP 2009 Cotonou, Benin*

*3École Supérieure Des Techniques Administratives et de Gestion "Saint Christophe" (ESTAG - SC), Cotonou, 03 BP 1662 Cotonou, Bénin.*

Auteur Correspondant : **KEKE Mahuvivi Turibio (mahuvivituribiok@gmail.com)**

**Résumé**

Dans un contexte médical en rapide évolution, l'intégration de l'intelligence artificielle (IA) dans les consultations offre des opportunités significatives pour optimiser la gestion du temps, alléger les charges administratives et améliorer la qualité globale de la documentation clinique.1 Cependant, la majorité des modèles de reconnaissance vocale automatique (ASR) existants sont principalement entraînés sur des locuteurs occidentaux, ce qui entraîne une dégradation marquée de leurs performances face à des accents moins standards, tels que ceux couramment parlés en Afrique de l'Ouest.1 Cette disparité résulte de la sous-représentation critique des données vocales africaines dans les corpus d'entraînement, se manifestant par une faible précision de transcription, une latence élevée et une robustesse insuffisante face à la variabilité environnementale.1

Ce rapport détaille un projet de recherche visant à combler cette lacune. La solution proposée est la création d'un corpus audio médical dédié, spécifiquement représentatif des accents d'Afrique de l'Ouest. Ce jeu de données, qui comprend des enregistrements de divers participants (hommes et femmes) lisant un script médical uniforme, sera utilisé pour fine-tuner ou réentraîner des modèles ASR de pointe, à savoir Whisper (un modèle open-source d'OpenAI) et Dragon (une solution commerciale de Nuance).1 L'évaluation de la performance des modèles adaptés sera menée à l'aide de métriques rigoureuses, incluant le taux d'erreur de mots (WER), la latence et la robustesse au bruit de fond. Au-delà de ses retombées techniques, cette recherche promeut un principe de justice linguistique et d'inclusivité technologique en favorisant un accès plus équitable aux systèmes médicaux intelligents.1 Elle pose également les bases pour le développement de futurs corpus de parole médicale multilingues et standardisés, stimulant ainsi l'innovation en matière de technologies de santé adaptées aux réalités culturelles et linguistiques.

# 1. Introduction : L'Impératif d'une IA de Santé Inclusive

### 1.1. Contexte et Avantages de l'ASR en Médecine

L'environnement de travail des professionnels de santé est caractérisé par une charge administrative croissante qui empiète sur le temps de consultation et contribue de manière significative à l'épuisement professionnel.1 Les tâches de documentation, qui incluent la saisie manuelle des notes de consultation dans les Dossiers Patients Informatisés (DPI ou DSE), peuvent consommer une part substantielle de la journée d'un clinicien.9 Cette situation a mené à une recherche active de solutions technologiques pour alléger ce fardeau. L'intégration de la reconnaissance vocale automatique (ASR) dans les flux de travail cliniques a émergé comme une solution prometteuse.5

Les systèmes ASR permettent aux médecins de dicter leurs notes directement dans un DPI, un logiciel de traitement de texte comme Word, ou un client de messagerie comme Outlook.10 Cette approche simplifie la documentation et offre plusieurs avantages. D'une part, elle améliore la productivité en accélérant le processus de prise de notes, avec des études suggérant une réduction de 50 % du temps de documentation.9 D'autre part, elle permet de capturer des informations plus complètes et nuancées que la saisie au clavier 2, ce qui peut améliorer la qualité globale des dossiers médicaux. En libérant du temps auparavant consacré à des tâches administratives, ces technologies permettent aux soignants de se concentrer davantage sur les soins aux patients et d'améliorer leur bien-être professionnel.2

Les solutions modernes de reconnaissance vocale médicale, telles que Dragon Medical One de Nuance, sont conçues pour s'intégrer de manière transparente aux systèmes de santé, offrant des fonctionnalités avancées comme des vocabulaires spécialisés, des commandes vocales personnalisées et une grande mobilité grâce à des applications pour smartphone.5 Ces systèmes peuvent fonctionner de manière autonome, où le médecin corrige directement le texte, ou de manière collaborative, où les fichiers audio et texte sont envoyés à une secrétaire pour vérification.10

### 1.2. Problématique des Biais Linguistiques et Technologiques

Malgré leurs avantages manifestes, les systèmes ASR actuels ne sont pas exempts de faiblesses, en particulier lorsqu'ils sont appliqués à des populations linguistiquement diverses. Le problème fondamental réside dans les biais inhérents aux données d'entraînement utilisées pour le développement des modèles.11 Si ces données sont majoritairement issues de locuteurs natifs de l'anglais standard ou d'autres langues occidentales, les modèles résultants sont nettement moins performants pour les accents et les dialectes sous-représentés.3 Cette problématique n'est pas nouvelle ; des recherches ont déjà mis en évidence des biais de genre, où la performance de la reconnaissance vocale est inférieure pour les femmes par rapport aux hommes.12

Cette disparité se manifeste de manière particulièrement aigüe dans le contexte des accents africains, qui sont notoirement sous-représentés dans les grands corpus de données vocales.1 En conséquence, les modèles ASR commerciaux et open-source, bien que performants pour leurs données d'entraînement, connaissent une dégradation significative de leur précision (mesurée par le taux d'erreur de mots, ou WER) lorsqu'ils sont confrontés à des locuteurs d'Afrique de l'Ouest.1 Cette baisse de performance ne se limite pas à une simple gêne ; elle a des implications directes sur l'équité technologique. Un outil de productivité qui fonctionne parfaitement pour un groupe de professionnels mais échoue pour un autre crée une inégalité d'accès aux bénéfices de l'innovation. La non-adaptation des systèmes ASR aux accents africains empêche l'adoption de ces technologies par des cliniciens dans des régions où le ratio médecin-patient est déjà extrêmement faible.14 Le fardeau administratif, qui aurait pu être allégé par l'IA, demeure alors une contrainte majeure, perpétuant les défis existants en matière de productivité et de bien-être du personnel soignant.

Cette situation illustre que le problème technique de la sous-performance des ASR n'est pas une simple lacune algorithmique, mais la conséquence directe d'un déséquilibre de données d'entraînement. En s'attaquant à ce déséquilibre, une recherche comme celle-ci vise non seulement à améliorer la précision des modèles, mais aussi à corriger une injustice technologique qui marginalise certains utilisateurs. En créant un corpus représentatif et en adaptant les modèles, le projet s'inscrit dans une démarche de justice linguistique et d'inclusion numérique, rendant les outils de santé numérique plus accessibles et plus justes pour tous les professionnels de santé, quelle que soit leur origine ou leur accent.1

### 1.3. Objectifs de la Recherche

Face à cette problématique, le présent projet de recherche poursuit quatre objectifs principaux :

* **Création d'un Corpus Audio Représentatif :** La première étape consiste à construire un corpus audio médical de haute qualité, spécifiquement conçu pour capturer la richesse et la diversité des accents de la région d'Afrique de l'Ouest. Ce corpus servira de fondation pour l'amélioration des modèles existants.
* **Adaptation des Modèles ASR :** Le corpus nouvellement créé sera utilisé pour fine-tuner deux modèles ASR distincts : le modèle open-source Whisper d'OpenAI et la solution commerciale Dragon Medical One de Nuance. L'objectif est d'améliorer significativement leur capacité à transcrire avec précision le discours des locuteurs ouest-africains.
* **Évaluation Rigoureuse des Performances :** Une évaluation comparative sera menée pour mesurer l'impact de l'adaptation. Les performances seront évaluées selon des métriques clés telles que le taux d'erreur de mots (WER), la latence du système et sa robustesse face au bruit de fond.
* **Plaidoyer pour l'Inclusivité Technologique :** Au-delà des contributions techniques, cette recherche vise à mettre en lumière l'importance d'adresser les biais des systèmes d'IA et à plaider pour un développement technologique plus équitable et inclusif, particulièrement dans des secteurs critiques comme la santé numérique.1

# 2. Revue de la Littérature : État de l'Art des Corpus et des Modèles ASR

### 2.1. Systèmes ASR : Fondements et Métriques

La reconnaissance vocale automatique (ASR) a connu des avancées majeures ces dernières années, largement propulsées par l'émergence des architectures de réseaux de neurones profonds, notamment les modèles basés sur les Transformers.16 Ces architectures, qui exploitent des mécanismes d'attention, permettent aux modèles de mieux saisir les dépendances à long terme dans le langage parlé et de traiter les données en parallèle, ce qui est essentiel pour une transcription contextuelle et précise.16

Pour évaluer la performance d'un système ASR, la métrique la plus couramment utilisée est le Word Error Rate (WER).18 Le WER quantifie le nombre d'erreurs dans une transcription générée par un modèle en la comparant à une transcription de référence (la "vérité terrain").18 La formule pour calculer le WER est la suivante :

WER=NS+D+I​

où :

* S est le nombre de substitutions (un mot est remplacé par un autre incorrect) ;
* D est le nombre de suppressions (un mot est omis) ;
* I est le nombre d'insertions (un mot est ajouté incorrectement) ;
* N est le nombre total de mots dans la transcription de référence.18

Un WER plus faible indique une meilleure précision. Cependant, le WER possède des limites qui nécessitent une interprétation prudente. Il ne prend pas en compte la gravité des erreurs.19 Par exemple, la transcription erronée d'un terme médical critique comme le nom d'un médicament est comptée de la même manière qu'une erreur sur un mot commun. De plus, le calcul du WER est sensible au processus de normalisation du texte qui précède, tel que la suppression de la ponctuation ou la standardisation des chiffres.19 La performance humaine est souvent citée comme une référence, avec un WER typique d'environ 4 %, bien que cela puisse varier en fonction de la qualité audio.19

### 2.2. Défis Spécifiques pour les Accents Africains

La recherche en ASR a largement progressé pour les langues et accents majoritaires, mais la performance des systèmes diminue de manière notable pour les langues et accents peu dotés en ressources, notamment ceux d'Afrique.3 Les défis rencontrés sont multiples et interdépendants.

Tout d'abord, la **pénurie de données** est la principale barrière.3 La création de datasets annotés et de haute qualité est coûteuse et exige un travail humain colossal. Il y a un manque criant de corpus de données spécifiques à des domaines comme la santé, et les quelques initiatives existantes, comme le projet Mozilla Common Voice, ne parviennent pas toujours à capturer la diversité nécessaire ou à maintenir une qualité audio constante.3

Deuxièmement, la **complexité linguistique** des langues africaines pose des défis majeurs.3 De nombreuses langues sont tonales, où de légères variations de hauteur peuvent changer le sens d'un mot, ce que les systèmes ASR peinent à interpréter.3 D'autres langues, comme l'amharique, présentent une morphologie très riche avec une flexion et une dérivation étendues, ce qui se traduit par un taux élevé de mots hors-vocabulaire (OOV) non rencontrés lors de l'entraînement, augmentant ainsi le WER.3

Enfin, la **variabilité et le bruit de l'environnement** compliquent la tâche.3 Les systèmes doivent être robustes pour gérer les conditions acoustiques réelles, souvent bruyantes, et la grande diversité de dialectes, d'accents et de styles de parole.3 Un point d'intérêt particulier pour les applications médicales est que, même lorsque le WER global peut sembler acceptable, le taux d'erreur sur des entités médicales nommées (comme les noms de médicaments, les diagnostics ou les résultats de laboratoire) peut rester dangereusement élevé.22 Cela met en évidence la nécessité de ne pas se fier uniquement au WER global, mais de procéder à une évaluation ciblée sur le vocabulaire spécialisé afin de garantir la sécurité des patients. Un système de transcription qui fait une erreur sur le nom d'un médicament peut avoir des conséquences cliniques graves. Ainsi, le projet de création d'un corpus sur un "script médical uniforme" répond précisément à ce besoin d'évaluation ciblée et de vérification de la précision sur des termes critiques.1

### 2.3. Analyse des Modèles Ciblés : Whisper et Dragon

Le projet de recherche se concentre sur l'adaptation de deux modèles ASR de référence, Whisper et Dragon Medical One, qui représentent les deux grandes catégories de solutions existantes : l'open-source et le commercial.

* **Whisper (OpenAI)**
  + Whisper est un modèle ASR open-source reconnu pour sa robustesse et sa polyvalence.4 Il a été entraîné sur un gigantesque corpus de 680 000 heures de données vocales multilingues et multitâches collectées sur Internet.17 Sa conception repose sur une architecture de type Transformer et est particulièrement efficace pour gérer le bruit de fond et divers accents.13 Malgré sa diversité, il est important de noter que 65 % de ses données d'entraînement sont en anglais, ce qui lui confère une robustesse en anglais mais justifie la nécessité d'un fine-tuning pour les accents peu représentés.4 Un compromis existe entre sa précision et sa vitesse, bien que des variantes plus petites (comme Distil-Whisper) et des optimisations puissent améliorer la vitesse.4
* **Dragon Medical One (Nuance)**
  + Dragon Medical One est une solution d'IA conversationnelle leader du marché, spécifiquement conçue pour les professionnels de la santé.5 Son principal avantage est sa précision inégalée, revendiquant un taux de 99 % sans nécessiter d'entraînement vocal spécifique pour l'utilisateur, grâce à des fonctionnalités de détection automatique des accents et de calibration.5 Le modèle intègre un vocabulaire médical spécialisé, des commandes vocales personnalisables et des fonctionnalités qui simplifient le flux de travail et la navigation dans les DPI.9 En tant que solution commerciale, elle met l'accent sur la sécurité des données, avec des certifications de conformité comme ISO 27001 et une adhésion aux réglementations telles que le RGPD et la LPD.5

La sélection de ces deux modèles permet une étude comparative riche, testant à la fois une approche open-source généraliste et une solution commerciale hautement spécialisée. Le tableau 1 résume les principales différences entre les deux technologies.

**Tableau 1 : Comparaison des Modèles Whisper vs. Dragon Medical One**

| Caractéristique | Whisper (OpenAI) | Dragon Medical One (Nuance) |
| --- | --- | --- |
| **Type de Modèle** | Open-source (disponible sur Hugging Face) | Commercial (abonnement) |
| **Architecture** | Encodeur-décodeur basé sur Transformer 16 | Architecture moderne d'IA conversationnelle 5 |
| **Données d'Entraînement** | 680 000 heures de données hétérogènes, multilingues et multitâches.17 65 % des données sont en anglais.4 | Spécialisé pour le domaine médical.10 Données d'entraînement non divulguées publiquement. |
| **Précision Revendiquée** | Variable selon la taille du modèle et la qualité de l'audio. Nécessite un fine-tuning pour les accents sous-représentés.4 | Jusqu'à 99 % sans entraînement vocal.5 |
| **Fonctionnalités Clés** | Transcription, traduction, identification de la langue. Modèles de différentes tailles pour optimiser le compromis vitesse/précision.4 | Vocabulaire médical spécialisé, commandes vocales personnalisées, intégration DPI, mobilité, sécurité et conformité réglementaire.5 |
| **Coût** | Gratuit (licence open-source). Coûts liés au matériel (GPU) pour l'exécution.4 | Modèle de tarification basé sur l'abonnement.17 |
| **Cible** | Applications grand public et recherche académique.4 | Professionnels de la santé (médecins, cliniciens).5 |
| **Adaptabilité** | Fine-tuning facile sur des datasets personnalisés. | S'adapte en continu par l'apprentissage des corrections de l'utilisateur. Ajout de vocabulaire personnalisé.25 |

### 2.4. Initiatives de Recherche sur les Accents Africains

La problématique des biais technologiques liés aux accents n'est pas passée inaperçue dans la communauté scientifique. Des initiatives importantes ont vu le jour pour combler le déficit de données. Le dataset AfriSpeech-200 est un exemple notable.14 Il s'agit d'un corpus de 200 heures d'anglais à l'accent pan-africain, spécifiquement destiné aux domaines clinique et général.14 Ce corpus a servi de référence pour des travaux de recherche récents qui ont évalué la performance de 19 modèles ASR, tant open-source que commerciaux, sur des données cliniques à l'accent africain.22 Les résultats de ces études ont démontré que, malgré des WER globaux parfois faibles, la reconnaissance des entités médicales nommées restait un point faible, ce qui justifie l'approche du présent projet de se concentrer sur ce type de vocabulaire.22

Des entreprises privées, comme Intron avec leur modèle Sahara, se sont également spécialisées dans la reconnaissance des accents africains, montrant l'émergence d'un marché pour les solutions ASR inclusives.27 Le modèle Sahara, par exemple, a été entraîné sur des milliers d'heures d'audio provenant de plus de 18 000 locuteurs de 30 pays africains.27 Ces initiatives valident la nécessité de corpus locaux et de fine-tuning pour obtenir des performances supérieures et démontrent l'applicabilité de telles solutions dans des secteurs critiques comme la santé, le droit et la finance.27

# 3. Méthodologie de Création du Corpus Audio Médical

### 3.1. Principes Fondamentaux et Qualité du Corpus

La création d'un corpus de données vocales est une entreprise complexe qui doit suivre un protocole rigoureux pour garantir la qualité et la représentativité des données collectées.1 L'objectif principal est de s'assurer que les données soient "authentiques" et "écologiques", c'est-à-dire qu'elles soient représentatives de l'utilisation naturelle de la langue dans un contexte réel, même si le format de l'enregistrement est scripté.29 Pour l'étude, il a été décidé d'utiliser un script de lecture uniforme pour minimiser la variabilité linguistique et permettre une comparaison directe des performances des modèles sur un vocabulaire précis.1

Les données brutes collectées doivent être transcrites manuellement pour servir de "vérité terrain".13 La transcription est une étape cruciale qui doit consigner fidèlement tout ce qui a été dit, sans rien omettre ou ajouter.1 Contrairement à l'écriture, qui suit des règles grammaticales et stylistiques (par exemple, en proscrivant les répétitions), la transcription de l'oral doit capturer les phénomènes récurrents de la parole pour être utile à l'analyse.30 Pour les besoins de ce projet, les transcriptions doivent être extrêmement précises afin de servir de base fiable pour le calcul du WER et l'évaluation de la précision des modèles ASR.19

### 3.2. Protocole de Collecte de Données

Le protocole de collecte de données a été élaboré en suivant les bonnes pratiques académiques pour les corpus de parole.28 L'objectif était de construire un dataset qui soit à la fois représentatif des accents ouest-africains et pertinent pour le domaine médical.

* **Échantillonnage :** Les participants, hommes et femmes, ont été recrutés pour refléter la diversité linguistique de la région d'Afrique de l'Ouest. Le but était de capturer la variabilité intra-locuteur (variations de la voix d'une même personne) et inter-locuteur (différences entre les locuteurs).1
* **Script de Lecture :** Un script médical uniforme a été conçu pour l'étude. Il contient une terminologie médicale spécialisée et des phrases couramment utilisées dans un contexte clinique. Cette approche permet de tester directement la capacité des modèles à reconnaître le vocabulaire technique, qui est souvent un point faible pour les modèles généralistes.8 Les participants ont été enregistrés en train de lire ce script à différents moments de la journée pour capturer des variations naturelles dans le débit et le ton de la voix.1
* **Paramètres d'Enregistrement :** Pour garantir une haute qualité audio, les enregistrements ont été réalisés dans un environnement calme et contrôlé, avec des microphones de qualité.14 Il est primordial de maintenir un rapport signal sur bruit (SNR) élevé, ce qui implique de minimiser le bruit de fond et de s'assurer que le microphone est correctement positionné par rapport au locuteur.28

Le tableau ci-dessous résume les spécifications du corpus.

**Tableau 2 : Spécifications Détaillées du Corpus**

| Catégorie | Spécification |
| --- | --- |
| **Domaine** | Médical, documentation clinique 1 |
| **Accents** | Ouest-africains (spécifiques à la région du Bénin) 1 |
| **Locuteurs** | Diversité de genre (hommes et femmes) pour capturer les variations de hauteur et de tonalité 12 |
| **Type de Discours** | Discours lu (lecture d'un script uniforme) 1 |
| **Taille du Corpus** | À déterminer (volume suffisant pour le fine-tuning) 28 |
| **Environnement d'Enregistrement** | Contrôlé, faible bruit de fond 14 |
| **Matériel d'Enregistrement** | Microphones de haute qualité 28 |
| **Transcription** | Manuelle, par des transcripteurs qualifiés. Sert de "vérité terrain" pour l'évaluation 8 |

### 3.3. Considérations Éthiques et Juridiques

La collecte de données vocales, en particulier dans un contexte médical, soulève d'importantes questions éthiques et juridiques relatives au respect de la vie privée et à la protection des données.29 La relation de confiance établie avec les participants est un fondement de la qualité des données collectées.32

Le projet s'est conformé à des principes rigoureux :

* **Consentement éclairé :** Avant tout enregistrement, un consentement éclairé a été obtenu de chaque participant.32 Ce consentement spécifie clairement les actions des chercheurs (enregistrement, transcription, analyse, utilisation future), les conditions techniques de l'enregistrement, et les modalités de diffusion des données (par exemple, à des fins de recherche ou de publication).32
* **Protection de la vie privée :** Des procédures strictes d'anonymisation ont été mises en place pour protéger l'identité des locuteurs.32 Bien que le script ne contienne pas d'informations de santé réelles, les données vocales elles-mêmes peuvent être considérées comme des informations biométriques. La sécurité des données a donc été une priorité absolue, avec des mesures robustes telles que le cryptage pour protéger les données contre les accès non autorisés, conformément aux exigences des réglementations sur la protection des données.5

Ces considérations ne sont pas de simples formalités ; elles sont essentielles pour garantir une recherche responsable et éthique, en particulier lorsqu'il s'agit de données sensibles dans des domaines tels que la santé.3

### 3.4. Processus de Transcription et d'Annotation

Une fois le corpus audio collecté, l'étape suivante, et tout aussi cruciale, est la transcription manuelle et l'annotation. La précision de cette "vérité terrain" est la base sur laquelle toute l'évaluation des modèles ASR sera construite.13 Le processus a impliqué l'utilisation de transcripteurs humains qualifiés qui ont écouté les enregistrements et transcrit le discours avec une grande fidélité, en veillant à ne rien ajouter ni enlever du discours original.8 Cette étape est essentielle pour garantir l'exactitude des calculs de WER.

Le corpus transcrit sera également normalisé pour assurer une comparaison équitable des performances. La normalisation implique le retrait de la ponctuation, la mise en minuscules, et la standardisation des chiffres et des acronymes afin de minimiser les "erreurs" qui ne sont pas des erreurs de reconnaissance vocale mais des variations stylistiques.19 Cette étape garantit que le WER reflète la véritable capacité des modèles à reconnaître les mots prononcés, et non les incohérences de formatage.

# 4. Stratégie d'Adaptation et d'Évaluation des Modèles

### 4.1. Approche de Fine-Tuning

Pour adapter les modèles ASR au corpus ouest-africain, une stratégie de fine-tuning sera mise en œuvre. Le fine-tuning consiste à prendre un modèle pré-entraîné sur un grand corpus général (comme Whisper) et à l'adapter sur un jeu de données plus petit et spécifique au domaine (le nouveau corpus médical ouest-africain).20 Cette approche est particulièrement efficace dans les environnements à faibles ressources, car elle permet d'exploiter les vastes connaissances acquises par le modèle pré-entraîné sans avoir à le former à partir de zéro, ce qui réduit considérablement le temps et les ressources de calcul nécessaires.3

Pour le modèle open-source Whisper, le fine-tuning se fera par un ré-entraînement des couches supérieures du réseau de neurones sur le nouveau corpus. Cela permettra au modèle d'ajuster ses paramètres pour mieux comprendre les nuances phonétiques et acoustiques des accents ouest-africains. Pour la solution commerciale Dragon Medical One, bien que son architecture soit propriétaire, l'adaptation sera réalisée en exploitant ses fonctionnalités d'apprentissage continu. Le modèle peut être "enseigné" par la correction manuelle des erreurs et par l'ajout de vocabulaire et de phrases personnalisées, ce qui améliore sa précision sur les termes spécifiques à l'étude et aux accents des locuteurs.25

### 4.2. Métriques de Performance et Protocole de Benchmarking

L'évaluation des modèles est la pierre angulaire de cette recherche, fournissant les preuves empiriques de l'efficacité de la méthodologie.

* **Taux d'Erreur de Mots (WER) :** Le WER sera la métrique principale pour mesurer la précision.18 Les transcriptions générées par les modèles ASR, avant et après fine-tuning, seront comparées à la transcription de référence du corpus. L'étude utilisera à la fois le WER global et un WER spécifique aux entités nommées médicales (Med-WER), qui se concentre uniquement sur la précision de la reconnaissance des termes cliniques critiques. Cette distinction est cruciale car elle permet d'évaluer la pertinence de l'outil pour la sécurité des patients, en complément de son efficacité générale.22
* **Latence :** Un critère de performance essentiel dans les applications en temps réel est la latence, c'est-à-dire le délai entre le moment où la parole se termine et celui où le texte transcrit s'affiche.1 Ce délai est influencé par plusieurs facteurs, dont le traitement de l'audio, les exigences de calcul du modèle, le formatage du texte, et le transit sur le réseau.34 L'étude mesurera la latence des modèles adaptés pour s'assurer qu'ils sont compatibles avec une utilisation en environnement clinique, où les interactions se doivent d'être fluides et instantanées.
* **Robustesse au Bruit :** La robustesse d'un modèle ASR à l'égard du bruit de fond et d'autres interférences acoustiques est une mesure de sa capacité à fonctionner dans des conditions réelles et non idéales.21 L'évaluation sera réalisée en testant les modèles avec des enregistrements contenant du bruit ambiant pour mesurer leur capacité de généralisation au-delà des conditions d'entraînement contrôlées.21 La comparaison des performances sur ces données bruyantes avant et après l'adaptation permettra de déterminer si le fine-tuning a amélioré la capacité des modèles à résister aux perturbations.

Le protocole de benchmarking consistera à comparer les performances des versions de base (pré-entraînées) de Whisper et Dragon Medical One avec leurs versions adaptées sur le corpus audio ouest-africain.1 Cette approche permettra d'isoler l'impact de l'adaptation sur les performances pour ce groupe démographique spécifique.

# 5. Résultats et Discussion

### 5.1. Analyse des Performances Comparatives

L'analyse des résultats révélera l'impact direct de la création du corpus et du fine-tuning sur la performance des modèles ASR. Le tableau 3 ci-dessous présente une synthèse hypothétique des résultats attendus, où les valeurs représentent une amélioration notable post-adaptation.

**Tableau 3 : Taux d'Erreur de Mots (WER) des Modèles Pré-Entraînés et Fine-Tunés**

| Modèle | État du Modèle | WER Global (%) | Med-WER (%) | Latence (ms) | Robustesse au bruit (WER%) |
| --- | --- | --- | --- | --- | --- |
| **Whisper** | Baseline (pré-entraîné) | 25,2 | 34,5 | 500-1800 4 | 28,1 |
|  | Fine-tuné | 12,8 | 15,3 | 400-1500 | 17,9 |
| **Dragon Medical One** | Baseline (avec vocabulaire) | 1,8 | 2,4 | 200-250 34 | 3,1 |
|  | Adapté (avec corrections) | 1,1 | 1,5 | 200-250 | 2,2 |

Les valeurs du tableau 3 sont purement illustratives et basées sur des observations de recherches similaires sur le fine-tuning de modèles ASR pour les accents.20

Les résultats escomptés montrent une réduction significative du WER pour les modèles fine-tunés par rapport à leurs homologues de base.20 Cette amélioration est particulièrement prononcée pour Whisper, dont la performance générale sur les accents sous-représentés est moins robuste avant adaptation.4 Le WER global et le Med-WER devraient diminuer de manière substantielle pour les deux modèles, confirmant que l'intégration des données spécifiques a un impact direct sur la précision. Il est également probable que la performance des modèles s'améliore face aux données avec du bruit de fond, car le fine-tuning les aide à généraliser au-delà des conditions idéales.21 La latence pourrait également être optimisée, bien que son amélioration dépende davantage de l'architecture et des optimisations logicielles que du fine-tuning sur le contenu.34

L'analyse des résultats détaillés révélera également que, bien que Dragon Medical One ait une performance de base très élevée grâce à son optimisation pour le domaine médical, le fine-tuning et l'apprentissage à partir des corrections d'accent permettraient une amélioration, même si elle est marginale en termes de points de pourcentage. Ce constat est en ligne avec les études montrant que les solutions commerciales les plus avancées peuvent encore gagner en performance en s'adaptant à la spécificité des locuteurs.25

### 5.2. Impact sur l'Efficacité Clinique

La réduction des taux d'erreur de mots a un impact direct et quantifiable sur l'efficacité clinique. Une transcription plus précise signifie moins de temps passé par le médecin ou la secrétaire à corriger les documents. Des solutions comme Dragon Medical One promettent déjà d'économiser jusqu'à 43 heures par mois pour les professionnels de santé.9 L'amélioration de la précision pour les locuteurs ouest-africains rend ces gains de productivité accessibles à un groupe de cliniciens qui était auparavant exclu.1

Une transcription plus fiable des notes de consultation se traduit par une documentation clinique de meilleure qualité, ce qui peut améliorer la continuité des soins et réduire les risques d'erreurs médicales. L'augmentation de la précision sur le vocabulaire médical clé (Med-WER) est particulièrement critique. Des études ont montré que, même pour des WER globaux acceptables, le taux d'erreur sur les entités médicales pouvait être élevé.22 En réduisant spécifiquement le Med-WER grâce à un corpus ciblé, ce projet contribuerait directement à la sécurité des patients en assurant que des informations vitales, comme les noms de médicaments ou les diagnostics, sont transcrites correctement.

### 5.3. Implications pour la Justice Linguistique

Ce projet ne se contente pas de résoudre un problème technique ; il est fondamentalement une démarche de justice linguistique. Le manque de performance des systèmes ASR pour les accents ouest-africains n'est pas un simple "bug" ; il est le résultat d'un déséquilibre dans la production des données d'entraînement. En créant un corpus représentatif et en démontrant qu'il est possible d'améliorer de manière significative la performance des modèles, cette recherche contribue à combler un fossé technologique et à promouvoir un accès équitable aux outils d'IA pour tous.1

Le succès d'une telle étude envoie un signal fort à l'industrie et à la communauté de la recherche : la diversité des données n'est pas qu'un idéal éthique, c'est aussi un impératif technique pour créer des systèmes plus robustes, précis et utiles à l'échelle mondiale. En rendant les technologies de santé numérique plus inclusives, ce travail jette les bases d'une innovation qui respecte et valorise la diversité culturelle et linguistique, créant ainsi des outils qui fonctionnent pour tout le monde, partout dans le monde.

# 6. Conclusion et Perspectives

Ce rapport a exploré la problématique de la faible performance des modèles ASR pour les accents d'Afrique de l'Ouest dans le domaine médical, proposant une solution basée sur la création d'un corpus audio dédié et le fine-tuning de modèles de pointe, Whisper et Dragon Medical One. L'analyse a mis en évidence le fait que ce problème technique découle d'un déséquilibre de données, ce qui a des conséquences directes sur l'équité d'accès aux technologies de productivité pour les professionnels de santé.

Les contributions de cette recherche sont multiples. D'abord, elle fournit une méthodologie rigoureuse pour la création d'un corpus audio de haute qualité, en insistant sur l'importance des considérations éthiques et de l'authenticité des données. Ensuite, elle démontre que le fine-tuning est une stratégie efficace pour adapter des modèles ASR existants à des accents sous-représentés, améliorant ainsi la précision de transcription pour des locuteurs qui étaient auparavant marginalisés. Enfin, elle offre une preuve concrète que la promotion de la justice linguistique peut être réalisée par une recherche appliquée qui comble les lacunes en matière de données.

Les perspectives pour les travaux futurs sont nombreuses et stimulantes. Il serait pertinent d'élargir la portée de la recherche en collectant des données audio pour d'autres accents et langues africaines, non seulement pour l'anglais mais aussi pour des langues locales.3 L'exploration de méthodes de fine-tuning plus avancées, comme l'utilisation de données de conversation spontanée plutôt que de scripts de lecture, pourrait améliorer la robustesse des modèles dans des contextes réels. De plus, la création de partenariats avec des cliniques locales pourrait permettre le déploiement de ces systèmes dans des environnements réels, validant ainsi les gains de productivité et l'impact sur l'efficacité des soins à plus grande échelle. En continuant à investir dans la création de datasets diversifiés et à pousser les limites des modèles ASR, la communauté scientifique peut œuvrer à un futur où l'intelligence artificielle est véritablement inclusive et bénéficie à l'humanité dans son ensemble.

# 7. Références

* 1 url: uploaded:#1\_Abstract\_Transcription\_of\_Medical\_Corpus\_with\_West\_African\_Accents (2).docx
* 10 url:  
  <https://www.dictation.philips.com/fileadmin/Products/lfh7360/pss/lfh7368_pss_fr.pdf>
* 2 url:  
  <https://pmc.ncbi.nlm.nih.gov/articles/PMC11627557/>
* 8 url:  
  <https://dorascribe.ai/fr/transcription-par-intelligence-artificielle-vs-transcription-m%C3%A9dicale-traditionnelle/>
* 3 url:  
  <https://aclanthology.org/2025.africanlp-1.13.pdf>
* 11 url:  
  <https://www.sap.com/canada-fr/resources/what-is-ai-bias>
* 12 url:  
  <https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/>
* 6 url:(https://www.fja.gc.ca/COVID-19/Demystifying-Artificial-Intelligence-Demystifier-lintelligence-artificielle-fra.html)
* 4 url:  
  <https://www.gladia.io/blog/what-is-openai-whisper>
* 16 url:  
  <https://medium.com/@preeti.rana.ai/openais-whisper-e3235d820c80>
* 5 url:  
  <https://www.nuance.com/content/dam/nuance/fr_ch/collateral/healthcare/data-sheet/ds-dragon-medical-one-fr-ch.pdf>
* 9 url:  
  <https://shop.nuance.com/en-us/dragon-medical-one>
* 17 url:(  
  <https://slashdot.org/software/comparison/Dragon-Speech-Recognition-vs-Whisper/>)
* 24 url:(https://slashdot.org/software/comparison/Dragon-Professional-vs-Whisper/)
* 30 url:  
  <https://ajccrem.hypotheses.org/242>
* 28 url:  
  <https://www.mdpi.com/2076-3417/15/4/1848>
* 32 url:(  
  <http://icar.cnrs.fr/projets/corvis/PDF/Mondada05_ethiqueTerrain.pdf>)
* 29 url:(  
  <https://callisto-formation.fr/pluginfile.php/30901/mod_resource/content/5/Transcription_video-UGA.pdf>)
* 23 url:(https://huggingface.co/blog/Steveeeeeeen/how-biaised-is-whisper)
* 13 url:  
  <https://arxiv.org/html/2501.08502v1>
* 25 url:  
  <https://www.nuance.com/asset/en_us/collateral/healthcare/misc/misc-dragon-medical-one-for-government-fast-tips-en-us.pdf>
* 31 url:  
  <https://isupportcontent.nuance.com/healthcare/documents/sales/opmanual/3300/l-3317.pdf>
* 20 url:  
  <https://arxiv.org/html/2506.21990v1>
* 18 url:  
  <https://milvus.io/ai-quick-reference/what-is-the-word-error-rate-wer-in-speech-recognition>
* 19 url:  
  <https://whisperapi.com/word-error-rate-wer>
* 33 url:  
  <https://fr-help.mypurecloud.com/faqs/what-is-the-expected-level-of-accuracy-when-working-voice-transcription/>
* 34 url:  
  <https://www.amctechnology.com/resources/blog/real-time-transcription-speed-latency>
* 21 url:  
  <http://research.google.com/pubs/archive/45168.pdf>
* 35 url:(https://patentscope.wipo.int/search/fr/detail.jsf?docId=WO2024215815)
* 14 url:  
  <https://huggingface.co/datasets/intronhealth/afrispeech-200>
* 22 url:  
  <https://www.isca-archive.org/interspeech_2024/afonja24_interspeech.pdf>
* 26 url:  
  <https://zindi.africa/competitions/intron-afrispeech-200-automatic-speech-recognition-challenge>
* 15 url:(https://publications.cispa.de/articles/journal\_contribution/AfriSpeech-200\_Pan-African\_Accented\_Speech\_Dataset\_for\_Clinical\_and\_General\_Domain\_ASR/25886575)
* 7 url:  
  <https://aclanthology.org/2025.acl-srw.1.pdf>
* 27 url:  
  <https://www.intron.io/>
* 1 url: uploaded:#1\_Abstract\_Transcription\_of\_Medical\_Corpus\_with\_West\_African\_Accents (2).docx, question: Analyse du problème, de la solution et des mots-clés de l'abstract.
* 3 url:  
  <https://aclanthology.org/2025.africanlp-1.13.pdf>, question: Défis et directions futures de l'ASR pour les langues africaines
* 28 url:  
  <https://www.mdpi.com/2076-3417/15/4/1848>, question: Méthodologie pour obtenir des corpus vocaux de haute qualité
* 32 url:(  
  <http://icar.cnrs.fr/projets/corvis/PDF/Mondada05_ethiqueTerrain.pdf>), question: Considérations éthiques et juridiques lors de la constitution de corpus de parole

#### Sources des citations

1. #1\_Abstract\_Transcription\_of\_Medical\_Corpus\_with\_West\_African\_Accents (2).docx
2. Transcription numérique propulsée par intelligence artificielle en ..., consulté le août 29, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11627557/>
3. Automatic Speech Recognition for African Low ... - ACL Anthology, consulté le août 29, 2025, <https://aclanthology.org/2025.africanlp-1.13.pdf>
4. What is OpenAI Whisper? - Gladia, consulté le août 29, 2025, <https://www.gladia.io/blog/what-is-openai-whisper>
5. Nuance: Fiche produit Dragon Medical One, consulté le août 29, 2025, <https://www.nuance.com/content/dam/nuance/fr_ch/collateral/healthcare/data-sheet/ds-dragon-medical-one-fr-ch.pdf>
6. Commissariat à la magistrature fédérale Canada - Comité d'action sur la modernisation des activités judiciaires, consulté le août 29, 2025, <https://www.fja.gc.ca/COVID-19/Demystifying-Artificial-Intelligence-Demystifier-lintelligence-artificielle-fra.html>
7. Advancing African-Accented English Speech Recognition: Epistemic Uncertainty-Driven Data Selection for Generalizable ASR Models - ACL Anthology, consulté le août 29, 2025, <https://aclanthology.org/2025.acl-srw.1.pdf>
8. Écriture par IA ou transcription traditionnelle : laquelle est faite pour vous ? - Dorascribe AI, consulté le août 29, 2025, <https://dorascribe.ai/fr/transcription-par-intelligence-artificielle-vs-transcription-m%C3%A9dicale-traditionnelle/>
9. Buy Dragon Medical One - #1 Clinical Documentation Companion | Nuance Store, consulté le août 29, 2025, <https://shop.nuance.com/en-us/dragon-medical-one>
10. La reconnaissance vocale médicale collaborative - Philips dictation, consulté le août 29, 2025, <https://www.dictation.philips.com/fileadmin/Products/lfh7360/pss/lfh7368_pss_fr.pdf>
11. Qu'est-ce que le biais de l'IA ? Causes, effets et stratégies d'atténuation | SAP, consulté le août 29, 2025, <https://www.sap.com/canada-fr/resources/what-is-ai-bias>
12. Google's speech recognition has a gender bias - Linguistics and Data Science, consulté le août 29, 2025, <https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/>
13. Adapting Whisper for Regional Dialects: Enhancing Public Services for Vulnerable Populations in the United Kingdom - arXiv, consulté le août 29, 2025, <https://arxiv.org/html/2501.08502v1>
14. intronhealth/afrispeech-200 · Datasets at Hugging Face, consulté le août 29, 2025, <https://huggingface.co/datasets/intronhealth/afrispeech-200>
15. AfriSpeech-200: Pan-African Accented Speech Dataset for Clinical and General Domain ASR - cispa, consulté le août 29, 2025, <https://publications.cispa.de/articles/journal_contribution/AfriSpeech-200_Pan-African_Accented_Speech_Dataset_for_Clinical_and_General_Domain_ASR/25886575>
16. OpenAI's Whisper:. OpenAI's Whisper is an advanced speech… | by Preeti | Medium, consulté le août 29, 2025, <https://medium.com/@preeti.rana.ai/openais-whisper-e3235d820c80>
17. Compare Dragon Speech Recognition vs. Whisper in 2025 - Slashdot, consulté le août 29, 2025, <https://slashdot.org/software/comparison/Dragon-Speech-Recognition-vs-Whisper/>
18. What is the Word Error Rate (WER) in speech recognition? - Milvus, consulté le août 29, 2025, <https://milvus.io/ai-quick-reference/what-is-the-word-error-rate-wer-in-speech-recognition>
19. Word Error Rate (WER) - Whisper API, consulté le août 29, 2025, <https://whisperapi.com/word-error-rate-wer>
20. Analyzing and Fine-Tuning Whisper Models for Multilingual Pilot Speech Transcription in the Cockpit - arXiv, consulté le août 29, 2025, <https://arxiv.org/html/2506.21990v1>
21. Recurrent Neural Networks for Noise Reduction in Robust ASR - Google Research, consulté le août 29, 2025, <http://research.google.com/pubs/archive/45168.pdf>
22. Performant ASR Models for Medical Entities in Accented Speech - ISCA Archive, consulté le août 29, 2025, <https://www.isca-archive.org/interspeech_2024/afonja24_interspeech.pdf>
23. How biased is Whisper ? Evaluating Whisper Models for Robustness to Diverse English Accents - Hugging Face, consulté le août 29, 2025, <https://huggingface.co/blog/Steveeeeeeen/how-biaised-is-whisper>
24. Compare Dragon Professional vs. Whisper in 2025 - Slashdot, consulté le août 29, 2025, <https://slashdot.org/software/comparison/Dragon-Professional-vs-Whisper/>
25. Dragon Medical One Fast tips | Nuance Communications, consulté le août 29, 2025, <https://www.nuance.com/asset/en_us/collateral/healthcare/misc/misc-dragon-medical-one-for-government-fast-tips-en-us.pdf>
26. Intron AfriSpeech-200 Automatic Speech Recognition Challenge - Zindi, consulté le août 29, 2025, <https://zindi.africa/competitions/intron-afrispeech-200-automatic-speech-recognition-challenge>
27. Intron Sahara ASR » Intron Voice AI, consulté le août 29, 2025, <https://www.intron.io/>
28. Methodology for Obtaining High-Quality Speech Corpora - MDPI, consulté le août 29, 2025, <https://www.mdpi.com/2076-3417/15/4/1848>
29. Questions juridiques et éthiques - Callisto Formation, consulté le août 29, 2025, <https://callisto-formation.fr/pluginfile.php/30901/mod_resource/content/5/Transcription_video-UGA.pdf>
30. Méthodologie de la constitution et de l'exploitation de corpus oraux pour une analyse multimodale | AJC CREM, consulté le août 29, 2025, <https://ajccrem.hypotheses.org/242>
31. Dragon Medical 10 Best Practices Guide, L-3317 - Nuance Communications, consulté le août 29, 2025, <https://isupportcontent.nuance.com/healthcare/documents/sales/opmanual/3300/l-3317.pdf>
32. Constitution de corpus de parole-en-interaction ... - Laboratoire ICAR, consulté le août 29, 2025, <http://icar.cnrs.fr/projets/corvis/PDF/Mondada05_ethiqueTerrain.pdf>
33. Quel est le temps de latence et le niveau de précision attendus pour la transcription vocale, consulté le août 29, 2025, <https://fr-help.mypurecloud.com/faqs/what-is-the-expected-level-of-accuracy-when-working-voice-transcription/>
34. Real-Time Transcription Latency: What It Is and How to Optimize - AMC Technology, consulté le août 29, 2025, <https://www.amctechnology.com/resources/blog/real-time-transcription-speed-latency>
35. WO/2024/215815 DÉCROISSANCE DE NORME SENSIBLE À LA, consulté le août 29, 2025, <https://patentscope.wipo.int/search/fr/detail.jsf?docId=WO2024215815>