Nos últimos anos, o reconhecimento automático da fala (*ASR*) tem tido um avanço considerável, proporcionando novas oportunidades em diversas áreas de aplicação, nomeadamente em aplicações direcionadas a crianças. A tecnologia *ASR* é promissora para interacções inovadoras entre humanos e máquinas através da voz, tutores de leitura automática e assistentes de patologia da fala, entre outras aplicações. No entanto, os desafios únicos colocados pela fala das crianças dificultam o desempenho dos actuais sistemas ASR concebidos para adultos. Por conseguinte, há uma necessidade urgente de soluções feitas à medida que respondam à fala das crianças. A investigação apresentada nesta tese dá passos importantes nesta direção, contribuindo com conhecimentos e metodologias valiosos para o desenvolvimento de sistemas de ASR optimizados para crianças, fazendo avançar esta área de investigação e abrindo novas possibilidades para a utilização da tecnologia de *ASR* para crianças.

Na fase inicial desta tese, estabelecemos as bases da nossa investigação, fornecendo uma visão geral do contexto e dos desafios consideráveis associados ao reconhecimento da fala das crianças. O principal desafio caracteriza-se pela significativa variabilidade presente nas componentes acústicas e linguísticas da fala das crianças, agravada pela variabilidade intra e inter-falantes. A resposta a estes desafios multifacetados inerentes ao reconhecimento do discurso infantil exige o acesso a um volume substancial de dados, o que constitui um segundo grande desafio. Enquanto os conjuntos de dados para a fala dos adultos são cada vez mais abundantes, os corpora da fala das crianças continuam a ser raros e geralmente limitados em tamanho. Nesta parte preliminar da tese, analisamos as diferentes metodologias empregues na literatura para abordar os diferentes desafios associados à *ASR* infantil. Através de uma análise aprofundada da literatura, identificamos as abordagens mais promissoras, que servem de base para o desenvolvimento desta tese. Além disso, fornecemos uma compilação não exaustiva de corpora de fala infantil documentados na literatura, representando a coleção mais abrangente disponível até à data. Esta compilação serve como um recurso valioso para os investigadores, oferecendo uma visão dos conjuntos de dados existentes para investigação e experimentação adicionais no contexto do reconhecimento do discurso infantil.

Na segunda parte da tese, centrámo-nos na exploração de um sistema híbrido de *ASR* baseado em *Hidden Markov Models*, especificamente concebido para o reconhecimento da fala de crianças. O foco foi tanto a língua inglesa como uma língua com poucos recursos, nomeadamente o português europeu. Os nossos esforços centraram-se na avaliação exaustiva de várias abordagens de transferência de conhecimentos, a fim de avaliar a sua eficácia no domínio do *ASR* infantil. Das várias abordagens avaliadas, a aprendizagem por transferência surgiu como a técnica mais promissora para sistemas dedicados exclusivamente ao reconhecimento da fala infantil. A aprendizagem multitarefa, por outro lado, revelou-se eficaz em cenários em que o sistema tinha de reconhecer simultaneamente a fala de crianças e de adultos. Além disso, introduzimos uma nova abordagem, designada "aprendizagem por transferência multilingue", que combina elementos das metodologias de aprendizagem multitarefa e de transferência. Os nossos resultados demonstraram a eficácia do treino de um sistema *ASR* infantil multilingue como uma inicialização superior para a subsequente aprendizagem por transferência num conjunto de dados infantis alvo, particularmente em contextos de poucos recursos. Esta abordagem revelou-se fundamental para atenuar os desafios associados à disponibilidade limitada de dados, abrindo caminho para sistemas de reconhecimento da fala infantil mais robustos e precisos em diversos contextos linguísticos.

Na fase subsequente desta tese, o nosso foco passou a ser a exploração do paradigma e*nd-to-end*, com o objetivo de fazer avançar as actuais abordagens no *ASR* infantil. Partindo da abordagem convencional de transferência de aprendizagem sobre todo o modelo, propusemos uma estratégia de avaliação mais nuançada. A nossa investigação revelou o papel fundamental do codificador no processo de afinação da *ASR* infantil de *end-to-end*. Esta descoberta vai de encontro à ideia de que de que, no contexto da fala das crianças, a variabilidade acústica supera significativamente os factores linguísticos, contribuindo para a degradação da precisão do reconhecimento. Para além disso, os nossos resultados evidenciaram a eficácia a eficácia da segmentação de camadas superiores, situadas mais perto da saída do codificador. Estes conhecimentos oferecem recomendações valiosas para otimizar o desenvolvimento de modelos *ASR* para crianças através da aprendizagem por transferência. Nesta secção da tese, também introduzimos o novo conceito de "afinação parcial" para arquitecturas baseadas em Transformadores. As nossas descobertas indicaram que o *fine-tuning* de componentes específicos da rede superou a abordagem tradicional de ajuste fino de todo o modelo. Nomeadamente, observámos que ajustar especificamente a componente *Feed-Forward Network* produziu os melhores resultados. Esta abordagem inovadora é promissora para melhorar o desempenho dos sistemas *ASR* infantis, concentrando os esforços de afinação nos componentes de rede com maior impacto.

Em seguida, motivados pela necessidade de uma transferência de conhecimentos eficiente em termos de parâmetros, em particular em cenários com dados de treino limitados, investigámos a utilização de módulos *Adapters*. Estes módulos, compostos por duas camadas lineares integradas num modelo congelado pré-treinado, oferecem um mecanismo para a transferência de conhecimentos, mantendo os pesos e os conhecimentos encapsulados no modelo pré-treinado. A nossa investigação abrangeu a avaliação de várias configurações nas arquitecturas *Transformer* e *Conformer*. Entre as inúmeras configurações avaliadas, a configuração paralela, juntamente com a sua extensão *Conformer* conhecida como “*Two Parallel Adapters*”, emergiu como a melhor para transferir conhecimento para a fala das crianças. De forma notável, essas configurações superaram o desempenho do ajuste fino de todo o modelo, alcançando resultados superiores ao usar apenas 10% dos parâmetros envolvidos no *transfer learning* tradicional. Esta avaliação sublinhou a promessa dos *Adapters* no contexto do ASR de crianças, sugerindo o seu potencial para uma adaptação mais precisa. Para melhorar ainda mais a adaptabilidade, introduzimos um procedimento não supervisionado em que os enunciados foram agrupados utilizando o método *k-means* aplicado à respectiva incorporação do locutor. Este método é justificado pela ideia de que a fala com características acústicas comparáveis, tal como detectada pelo extrator de incorporação do locutor, seria melhor se fosse utilizado um adaptador treinado em padrões de fala semelhantes, por oposição a um adaptador infantil geral.

Expandindo a eficácia dos adaptadores para colmatar a lacuna entre os domínios de origem e de destino para a fala infantil, utilizámos os *Adapters* para melhorar o aumento de dados com dados imperfeitos para a ASR infantil. Especificamente, introduzimos o método inovador "*Double Way Adapter Tuning*", que utiliza dados *Text-to-Speech (TTS)* como aumento de dados. Reconhecendo que o discurso sintético apresenta muitas vezes incompatibilidades acústicas com o discurso real, era imperativo atenuar esta lacuna, um desafio que a nossa abordagem "*Double Way Adapter Tuning*" pretendia resolver. Este método consiste num procedimento de duas etapas: inicialmente, treinar os módulos do *Adapters* utilizando dados *TTS* imperfeitos, seguido de um *fine-tune* dos *Adapters* e dos pesos de todo o modelo utilizando uma combinação de dados sintéticos e reais. Nomeadamente, os dados foram submetidos a uma abordagem distinta de via dupla, com o discurso sintético a passar pelos *Adapters* enquanto o discurso real os contornava. A implementação da abordagem "*Double Way Adapter Tuning*" produziu melhorias significativas em relação à linha de base e às técnicas anteriores nas arquitecturas *Transformer* e *Conformer*, destacando a eficácia do nosso método. Além disso, alargámos a filtragem da incorporação do altifalante de dados sintéticos imperfeitos, incorporando *x-vectors* em vez de vectores. Isto envolveu a utilização da semelhança de cosseno entre a referência e os enunciados gerados para descartar enunciados que possam ter sido gerados incorretamente, melhorando assim a qualidade e a fiabilidade dos dados sintéticos utilizados.

Inspirando-nos nos diversos sucessos observados com os *Adapters*, abrangendo tanto a transferência de *Adapters* como a inovadora abordagem "*Double Way Adapter Tuning*", avaliámos diferentes metodologias alternativas presentes na literatura. As nossas conclusões sublinharam a eficácia duradoura dos adaptadores tradicionais como o método mais eficaz e eficiente em termos de parâmetros para melhorar o desempenho do ASR das crianças. Verificámos que existe um compromisso entre a precisão e a eficiência dos parâmetros. Embora alguns métodos apresentassem uma eficiência de parâmetros elevada, muitas vezes resultavam em resultados significativamente degradados. Por outro lado, outras abordagens, embora menos eficientes em termos de parâmetros, produziram resultados comparáveis ou mesmo superiores ao ajuste fino de todo o modelo. Para atenuar este compromisso, introduzimos uma nova abordagem que aproveita a redundância inerente aos componentes da rede de alimentação dos modelos baseados em transformadores. A nossa metodologia "*Shared-Adapters*", em que um único *Adapter* é partilhado por todas as camadas em vez de ser atribuído um por camada, demonstrou um desempenho notável, superando o ajuste fino de todo o modelo. Apesar de enfrentar uma degradação mínima dos resultados de ASR em comparação com os *Adapters* tradicionais, o nosso adaptador partilhado foi treinado com um número substancialmente menor de parâmetros do que qualquer abordagem anterior dentro desta gama de desempenho. Assim, a metodologia “*Shared-Adapters*” surge como uma excelente candidata para alcançar uma transferência de eficiência de parâmetros superior no contexto da *ASR* infantil, abrindo caminho para metodologias de adaptação mais eficazes.