From Voices to Data: Tools for Creating a Multitask Atypical Speech *Corpus* through Citizen Science

Caio Oliveira Di Gioia (Computer Science)
caiodigioia@usp.br
Department of Computing and Mathematics, FFCLRP, USP

Victor Hugo S. Lembor (Biomedical Informatics)
victorlembor@usp.br
Ribeirão Preto Medical School, University of São Paulo

Samira Fares (Speech Therapy)
samirafares@usp.br
Ribeirão Preto Medical School, University of São Paulo

Alessandra Alaniz Macedo (Advisor)
ale.alaniz@usp.br
Department of Computing and Mathematics, FFCLRP, USP

ABSTRACT

Speech data from people with atypical speech patterns remain underrepresented in research, limiting the development of effective assistive technologies for communication support. The SofiaFala Ecoa project addresses this gap by designing and deploying an accessible, citizen-science-driven platform for the collaborative collection of a multitask atypical speech *corpus*. This initiative combines a redesigned mobile application, a web-based recording portal, and interdisciplinary outreach activities to engage participants with speech disorders, their families, and healthcare professionals. The system enables the capture of speech in multiple tasks—such as reading, repetition, and spontaneous speech—while supporting multimodal data integration. Our objectives are threefold: (1) to establish a scalable and inclusive infrastructure for gathering speech data; (2) to expand the availability of publicly shareable *corpora* of atypical speech; and (3) to foster community participation in research on assistive speech technologies. The contributions of SofiaFala Ecoa include the creation of a pilot *corpus* covering multiple speech tasks, the implementation of tools and protocols to ensure ethical and secure data collection, and the demonstration of feasibility through initial engagement metrics. By bridging accessibility, inclusivity, and technological innovation, SofiaFala Ecoa paves the way for improved AI-based speech technologies that reflect the needs of people with speech disorders.

KEYWORDS

Atypical speech *corpus*, Data collection tools, Assistive technology

1 INTRODUÇÃO

A fala é um dos principais meios de comunicação humana, desempenhando papel central na expressão de pensamentos, emoções e necessidades dos seres humanos. Indivíduos com distúrbios de fala resultantes de condições como paralisia cerebral, distúrbios do espectro autista, doenças neuromusculares, traumatismos ou sequelas de acidentes vasculares cerebrais frequentemente enfrentam barreiras significativas para a comunicação verbal. Essas limitações impactam não apenas a interação social, mas também o acesso à educação, à autonomia e à qualidade de vida.

Avanços em Inteligência Artificial (IA) para reconhecimento e síntese de fala oferecem novas possibilidades de ferramentas assistivas, promovendo socialização e desenvolvimento cognitivo, afetivo, social e emocional [1–3]. Contudo, a escassez de *corpora* de fala atípica diversificados e de alta qualidade limita a robustez e generalização de modelos de aprendizado de máquina, especialmente em português, devido à menor disponibilidade de dados públicos e ao risco de viés.

O projeto *SofiaFala Ecoa* atua nesse contexto como iniciativa de ciência cidadã e extensão universitária, baseado em três pilares: ampliação e diversificação de dados de fala atípica por meio de plataforma adaptada; integração de múltiplas tarefas (repetição de frases, leitura, descrição de imagens e narrativas) para usos variados, como reconhecimento de fala, detecção de disfluências e predição de palavras; e engajamento comunitário, promovendo conscientização sobre barreiras de comunicação e colaboração aberta em pesquisa.

Este artigo apresenta ferramentas para criação e manipulação de um *corpus* multitarefa de fala atípica, combinando coleta participativa com abordagem interdisciplinar em fonoaudiologia e ciência da computação. A Seção 2 aborda trabalhos relacionados, a Seção 3 detalha a *EColeta* e a *EProcessa* (ferramentas de coleta e manipulação de áudios), a Seção 4 discute os resultados, e a Seção 5 apresenta as considerações finais.

2 TRABALHOS RELACIONADOS

Essa seção descreve trabalhos de pesquisa considerando *corpora* de fala atípica, tecnologias assistivas de fala, ciência cidadã na coleta de dados, e protocolos multitarefas e multimodais, uma vez que estão relacionados às ferramentas *EColeta* e *EProcessa*.

2.1 Corpora de fala atípica

Diversos trabalhos recentes têm desenvolvido conjuntos de dados (*corpora*) de fala patológica em contextos clínicos e educacionais.

Saz *et al.* descrevem o *corpus* Alborada-I3A contendo gravações de fala de 14 crianças com deficiência de fala e 232 controles típicos [4]. De modo semelhante, o *corpus* francês TYPALOC reúne 28 pacientes disártricos (três patologias) e 12 falantes saudáveis em condições de leitura e fala espontânea [5]. Em português europeu, há iniciativas menores (por exemplo, o *corpus* PHONODIS de Ramalho, com 26 crianças com transtornos fonológicos) [6].

Neumann *et al.* [7] apresentaram um *corpus* multimodal (áudio e vídeo) em inglês de 278 pacientes com esclerose lateral amiotrófica (ELA), contendo medidas acústicas, prosódicas e orofaciais. Bhat e Strik [8] destacam que *corpora* como o UA-Speech e o TORGO vêm sendo usados para treinar e avaliar sistemas de reconhecimento automático de fala (ASR) em inglês para usuários com disartria, mostrando ganhos de desempenho com bases de dados ampliadas.

In: V Concurso de Trabalhos de Iniciação Científica (CTIC 2025). Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web (CTIC'2025). Rio de Janeiro/RJ, Brasil. Porto Alegre: Brazilian Computer Society, 2025.

ISSN 2596-1683

No contexto do português brasileiro, destaca-se o projeto TaRSila [9], que reúne e valida *corpora* acadêmicos (NURC, ALIP, C-ORAL Brasil, entre outros) e estruturou o CORAA (*Corpus* de Áudios Anotados), com centenas de horas de fala espontânea. Apesar desses avanços, ainda faltam *corpora* colaborativos voltados à fala atípica. O projeto SofiaFala Ecoa busca preencher essa lacuna por meio de coletas participativas que integram a variedade linguística e prosódica.

2.2 Tecnologias assistivas de fala

Na área de tecnologia assistiva, Howarth *et al.* [10] descrevem o Voiceitt, um aplicativo voltado para usuários com disartria, projetado para otimizar a interface, reduzir o tempo de treinamento de voz e melhorar a usabilidade geral. Como esforço nacional relevante, Rissato e Macedo [11] apresentaram o SofiaFala, um aplicativo móvel inteligente de apoio à terapia da fala. Após cerca de dois anos, o SofiaFala foi adotado por aproximadamente 1.400 fonoaudiólogos, demonstrando sua relevância e impacto como ferramenta de apoio à terapia da fala. O projeto SofiaFala Ecoa visa complementar o SofiaFala, oferecendo dados reais de fala para treinamento de sistemas ASR e síntese de voz, além de apoiar pesquisas clínicas.

2.3 Ciência cidadã na coleta de dados

A ciência cidadã tem sido aplicada à coleta e anotação de dados de fala e saúde. Em linguagem inglesa, Semenzin *et al.* [12] testaram anotações via *crowdsourcing* (Zooniverse) em gravações espontâneas de crianças (incluindo crianças com Síndrome de Angelman), mostrando que a participação de cidadãos pode fornecer informações valiosas para pesquisas em linguagem. Embora nesse estudo a contribuição fosse feita via anotações, iniciativas como o SofiaFala Ecoa estendem esse conceito ao permitir que usuários doem suas próprias gravações de voz. O SofiaFala Ecoa inova ao envolver diretamente pessoas com distúrbios de fala (e suas redes de apoio) na gravação de novos dados de voz, ampliando o *corpus* e promovendo engajamento da comunidade.

2.4 Protocolos multitarefa e multimodais

Para enriquecer a coleta, alguns trabalhos adotam protocolos multimodais. Alhinti *et al.* [13] criaram o DEED, um banco de dados áudio-visual de fala disártrica em inglês, que inclui expressões emocionais registradas simultaneamente em áudio e vídeo. Neumann *et al.* [7] também integraram vídeo à captura de áudio para analisar correlações entre movimentos faciais e indicadores da progressão da ELA. Embora atualmente o SofiaFala Ecoa capture apenas áudio, há planos para incorporar vídeo facial e labial nas coletas, permitindo criar um *corpus* multimodal mais rico. Isso contribuirá tanto para estudos clínicos quanto para o desenvolvimento de tecnologias assistivas de fala mais robustas.

Em síntese, embora existam iniciativas relevantes em *corpora* de fala atípica, tecnologias assistivas, ciência cidadã e protocolos multimodais, ainda há carência de bases colaborativas abrangentes para o português brasileiro. O projeto SofiaFala Ecoa busca suprir essa lacuna por meio da coleta participativa de gravações de usuários com distúrbios de fala, integrando a variedade linguística e prosódica, e com planos de expandir para dados multimodais. Essa abordagem visa fortalecer pesquisas clínicas e impulsionar o reconhecimento e a síntese de fala mais inclusivos e robustos.

3 FERRAMENTAS SOFIAFALA ECOA

O projeto SofiaFala Ecoa possui duas ferramentas principais: a *EColeta* [1], dedicada à captura das doações de áudio, e a *EProcessa*, responsável pelo pré-processamento do material coletado.

3.1 EColeta

A ferramenta *EColeta* suporta a coleta de áudios com um *design* responsivo, projetado para otimizar a experiência de doação de áudio em múltiplos dispositivos por participantes com transtornos de fala. A *interface* usa elementos visuais familiares, como ícones de microfone para gravação, *stop* para interrupção, *play* para reprodução e lixeira para exclusão, visando a uma interação intuitiva.

A tela inicial (Ver Figura 1) apresenta os objetivos do projeto, a finalidade da coleta, as etapas do processo de doação e um contador que exibe o total de contribuições recebidas. Nessa mesma tela, o participante realiza um cadastro, fornecendo dados como e-mail, data de nascimento, gênero e a etiologia do transtorno de fala. Tais informações são coletadas exclusivamente para fins estatísticos, sendo a doação de áudio anônima. A política de anonimato é detalhada no Termo de Consentimento Livre e Esclarecido (TCLE) e no Termo de Consentimento de Uso de Imagem, Som e Voz (TCUISV), ambos disponíveis para consulta. Para participantes menores de 18 anos, é exigida a anuência ao Termo de Assentimento[2].

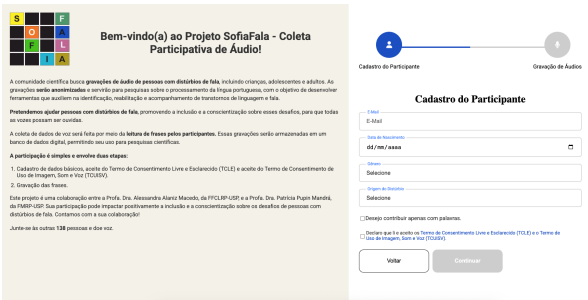

Figure 1: Tela inicial do site de coleta com contador de doações

Após o cadastro, o participante seleciona o tipo de contribuição — palavras ou frases — por meio de uma caixa de seleção (*checkbox*). A escolha direciona-o para o módulo correspondente.

O módulo de palavras (Ver Figura 2) exibe um estímulo multimodal, composto por uma imagem representativa, o texto da palavra e um botão que reproduz um áudio de referência. O participante pode gravar sua versão, ouvi-la e, se necessário, excluí-la caso a gravação não seja satisfatória.

De forma análoga, o módulo de frases apresenta o texto da frase e um botão para escutar o áudio de referência. As funcionalidades de gravação, reprodução e exclusão são idênticas às do módulo de palavras. Ao completar a gravação de dez áudios, o botão para

[1] https://sites.usp.br/sofiafala/coleta

[2] Esses termos e a pesquisa foram aprovados pelo Comitê de Ética em Pesquisa do HCFMRP-USP, CAEE n.95.853.018.0.0000.5440 pelo parecer de número 2.885.905.

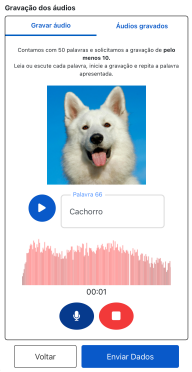

Figure 2: Tela de coleta dos áudios com uma imagem da palavra que está sendo doada (no exemplo, "Cachorro")

envio dos dados é habilitado. Após a submissão, uma mensagem de agradecimento é exibida, finalizando o processo de doação.

O projeto também prevê o uso de elementos de gamificação, como metas de doações e recompensas simbólicas, para estimular o engajamento. Além disso, investe em campanhas de divulgação nas redes sociais para alcançar potenciais doadores.

3.2 EProcessa

A ferramenta *EProcessa* é um *dashboard* administrativo, com *design* intuitivo e acessível aos membros do projeto mediante *login*. Foi projetado para gerenciar o pré-processamento dos áudios coletados, facilitando a avaliação e a classificação necessárias para a construção do *corpus*. O painel inicial exibe métricas, como o número total de doações, quantidade de áudios, idade média dos participantes e a distribuição por gênero. Um resumo quantitativo do status de classificação de todos os áudios também é apresentado neste painel. A plataforma permite a visualização de cada doação, possibilitando que os pesquisadores auditem os áudios individualmente e os classifiquem de acordo com as seguintes categorias: "Boa Qualidade", "Início Cortado", "Final Cortado", "Início Demorado", "Ruído", " Ruim", e "Não-verbal" (Ver Figura 3). Adicionalmente, a *EProcessa* oferece funcionalidades de gerenciamento, como a remoção de doações ou áudios individuais e a edição de metadados — por exemplo, a alteração do tipo de estímulo (palavra/frase) ou a correção da transcrição associada. A metodologia consistiu na coleta participativa e classificação manual, realizada por uma fonoaudióloga, contemplando diversidade linguística, prosódica e com diferentes faixas etárias. Cada registro reúne dados demográficos.

4 AVALIAÇÃO E RESULTADOS

A avaliação dos áudios na *EProcessa* contemplou tanto análises quantitativas, referente ao volume e a distribuição das classificações atribuídas aos áudios, quanto análises qualitativas, voltadas à classificação dos áudios, sendo realizada escuta de cada áudio e classificando com sua respectiva categoria. Por meio de avaliação especialista, foi possível avaliar as amostra, as quais eram acompanhadas dos seguintes metadados: idade média dos doadores e o sexo.

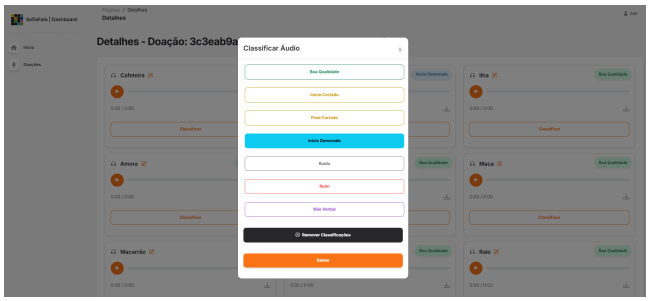

Figure 3: Tela de classificação de um áudio

4.1 Resultados Quantitativos

Até o momento, o projeto contabilizou a doação de 1.807 áudios, sendo alcançado em cerca de 1 ano de coleta. Um dos principais desafios foi a divulgação contínua e a dificuldade de atingir o público-alvo de forma efetiva. Todos foram submetidos a uma triagem inicial por uma fonoaudióloga para determinar a adequação do áudio para o *corpus*. A coleta foi aberta a participantes de ambos os sexos com um total de 138 doadores — 124 crianças com até 12 anos e os demais com mais de 12 —, sendo exigido doar no mínimo 10 áudios. Na doação, os participantes puderam optar por contribuir com frases ou palavras dependendo das diferentes condições ou preferência, conforme apresentado na Tabela 1. Dessa forma, cada participante contribuiu de acordo com seu perfil, colaborando para a diversidade linguística e prosódica do *corpus*. Independente da opção de doação, todos os conjuntos foram elaborados de modo a contemplar a totalidade dos fonemas do português brasileiro, assegurando a representatividade linguística.

Table 1: Opções de doação de áudios no *SofiaFala Ecoa*

Opção	Condição e Distúrbio de Fala (DF)
Frases infantis (até 10)	Idade ≤ 12 anos ou síndromes/transtornos
Frases gerais (até 50)	Idade > 12 anos e sem síndromes/transtornos
Palavras (até 100)	Qualquer participante com DF

No caso das frases, foram coletados 281 arquivos de áudio. Elas foram elaboradas para menores de 12 anos ou com diagnóstico de síndromes e maiores de 12 anos sem esse diagnóstico. Cada frase é composta por, no máximo, 8 palavras, para evitar construções excessivamente longas, que poderiam favorecer a aglutinação e comprometer a inteligibilidade da fala, em função das múltiplas coarticulações necessárias à produção de cada palavra. Essa divisão (idade e diagnóstico) visa incluir pessoas de diferentes categorias, contemplando suas aquisições e dificuldades fonológicas, que podem ou não estar associadas à sua condição.

Já o conjunto de palavras obteve a doação de 1.526 áudios, e essa opção de doação privilegiou termos de uso cotidiano, para atender todos os públicos, de modo a favorecer maior naturalidade na produção da fala e aumentar o engajamento dos doadores. Essa

estratégia contribuiu para que as gravações refletissem não apenas a diversidade fonética, mas também contextos comunicativos próximos à realidade dos falantes.

Além disso, visando compreender melhor as características da amostra, foi realizada a quantificação do sexo e da idade dos participantes. A análise revelou que a idade média dos doadores foi de 8 anos, com predominância do sexo feminino (77 participantes) em relação ao masculino (61 participantes). Para equilibrar a amostra quanto à idade e gênero, prevê-se ampliar a divulgação em diferentes canais e adotar estratégias de recrutamento direcionadas aos grupos sub-representados. Essas informações, aliadas aos registros qualitativos, permitem observar tendências de participação e possíveis variações na qualidade das gravações associadas a diferentes faixas etárias ou perfis demográficos.

4.2 Resultados Qualitativos

Dentre os áudios processados, 1.301 foram classificados como de "Boa Qualidade", representando a maior parte do material disponível. As demais classificações são apresentadas na Tabela 2.

Table 2: Classificações dos áudios avaliados no *EProcessa*

Categoria	Qtde	Observações
Boa Qualidade	1.301	Áudio adequado
Ruído	98	Áudio com falha de inteligibilidade
Ruim	462	Fala irreconhecível ou vazia
Início Demorado	65	Latência inicial da fala
Final Cortado	6	Perda do final da fala
Início Cortado	4	Perda do início da fala
Não verbais	43	Sons sem conteúdo verbal
Total de áudios	**1.807**	**Total áudios processados**

Nota: O total excede 1.807 áudios pois cada áudio pode ter múltiplas etiquetas.

A avaliação qualitativa teve como foco identificar inconsistências e verificar a clareza e aplicabilidade dos critérios de classificação adotados na *EProcessa*. O processo ocorreu por meio da escuta individual de cada áudio, considerando aspectos técnicos de qualidade e características linguísticas e comportamentais dos participantes. Durante o processo de coleta e análise, novas demandas surgiram. Um exemplo relevante foi a ocorrência de áudios provenientes de crianças não verbais, que produziam entonações e ritmos semelhantes à fala, mas sem articulação clara, demandando a criação de uma nova categoria de classificação ("Não verbais"), garantindo que tais registros fossem identificados e preservados para estudos futuros sobre padrões prosódicos e de desenvolvimento da comunicação. A Tabela 2 evidencia que a maior parte do material processado pela *EProcessa* tem qualidade satisfatória. As demais categorias revelam problemas de captação e execução da tarefa, que variam desde interferências sonoras até falhas na gravação, podendo variar conforme dispositivo e ambiente, sendo orientadas por instruções nas redes sociais; a coleta presencial reduziria essa influência, mas limitaria o alcance da diversidade de fala. A possibilidade de atribuição múltipla de etiquetas reforça a complexidade do processo de avaliação, permitindo caracterizar com maior precisão a coleta.

5 CONCLUSÃO

O SofiaFala Ecoa tem como objetivo central a criação de um *corpus* multitarefa de fala atípica em língua portuguesa, por meio de coleta colaborativa de pessoas com distúrbios de fala. As principais contribuições são: (1) uma infraestrutura de coleta online e presencial adaptada a públicos com limitações motoras e cognitivas, incorporando mecanismos de acessibilidade e consentimento informado inclusivo; (2) um protocolo multitarefa para gravação de fala atípica, visando a reutilização dos dados em pesquisa e criação de tecnologias assistivas; (3) a formação de um *corpus* de fala atípica anotado e disponibilizado para a comunidade científica sob licenças adequadas, fomentando a replicabilidade e a inovação; (4) relatos das experiências e desafios do processo de coleta colaborativa, incluindo aspectos éticos, técnicos e sociais, contribuindo para a construção de diretrizes em pesquisas participativas sobre fala.

Com esta iniciativa, espera-se disponibilizar recursos inéditos à comunidade, fortalecer o protagonismo de pessoas com distúrbios de fala, possibilitar soluções que as beneficie diretamente, e, futuramente, iniciar a captura de vídeos para criação de novas *corpora*.

REFERENCES

[1] K. Pedro and M. Chacon, "Softwares educativos para alunos com deficiência intelectual: estratégias utilizadas," *Rev. Br de Educação Especial*, vol. 19, no. 2, pp. 195–210, 2013.

[2] J. Carrer, E. B. Pizzolato, and C. Goyos, "Avaliação de software educativo com reconhecimento de fala em indivíduos com desenvolvimento normal e atraso de linguagem," *Rev. Brasileira de Informática na Educação*, vol. 17, no. 03, p. 67, 2009.

[3] D. de Souza, D. dos Santos, Nascimento, and E. Schlüzen, "Uso das tecnologias de informação e comunicação para pessoas com necessidades educacionais especiais como contribuição para inclusão social, educacional e digital," *Rev. Educação Especial*, pp. 25–36, 2005.

[4] O. Saz, E. Lleida, C. Vaquero, and W.-R. Rodríguez, "The alborada-I3A corpus of disordered speech," in *Proc. of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA), May 2010.

[5] C. Meunier, C. Fougeron, C. Fredouille, B. Bigi, L. Crevier-Buchman, E. Delais-Roussarie, L. Georgeton, A. Ghio, I. Laaridh, T. Legou, C. Pillot-Loiseau, and G. Pouchoulin, "The TYPALOC corpus: A collection of various dysarthric speech recordings in read and spontaneous styles," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 4658–4665. [Online]. Available: https://aclanthology.org/L16-1738/

[6] A. M. M. C. Ramalho, "Aquisição fonológica na criança: tradução e adaptação de um instrumento de avaliação interlinguístico para o português europeu," Ph.D. dissertation, Universidade de Évora, May 2018, orientadores: Maria João Freitas, Fernanda Gonçalves, Dina Caetano Alves. [Online]. Available: http://hdl.handle.net/10174/23564

[7] M. Neumann, H. Kothare, and V. Ramanarayanan, "Multimodal speech biomarkers for remote monitoring of ALS disease progression," *Comput Biol Med*, vol. 180, p. 108949, Aug. 2024.

[8] C. Bhat and H. Strik, "Speech technology for automatic recognition and assessment of dysarthric speech: An overview," *J Speech Lang Hear Res*, vol. 68, no. 2, pp. 547–577, Jan. 2025.

[9] C4AI - Centro de Inteligência Artificial da USP, "TaRSila," https://sites.google.com/view/tarsila-c4ai, acessado em: 18 ago. 2025.

[10] E. Howarth, G. Vabulas, S. Connolly, D. Green, and S. Smolley, "Developing accessible speech technology with users with dysarthric speech," *Assist Technol*, pp. 1–8, Mar. 2024.

[11] P. Rissato and A. Macedo, "Sofiafala: Software inteligente de apoio à fala," in *Anais Estendidos do XXVII Simpósio Brasileiro de Sistemas Multimídia e Web*, Porto Alegre, RS, Brasil, 2021, pp. 91–94.

[12] C. Semenzin, L. Hamrick, A. Seidl, B. L. Kelleher, and A. Cristia, "Describing vocalizations in young children: A big data approach through citizen science annotation," *J Speech Lang Hear Res*, vol. 64, no. 7, pp. 2401–2416, Jun. 2021.

[13] L. Alhinti, S. Cunningham, and H. Christensen, "The dysarthric expressed emotional database (DEED): An audio-visual database in british english," *PLoS One*, vol. 18, no. 8, p. e0287971, Aug. 2023.