

Pemanfaatan Teknologi *Speech-to-Text* untuk Penilaian Diri dalam Pengucapan Bahasa Inggris bagi Pembelajar EFL

Rini Puspasari¹, Maharani Nur Khafifah², Ulil Albab³

¹ SMK YPE Sampang Cilacap, ² SMK YPE Sampang Cilacap, ³ UNUGHA

Correspondence: rinipuspasari0408@email.com

Article Info	ABSTRACT
Article history: Received Jul 22 th , 2025 Revised Aug 18 th , 2025 Accepted Aug 18 th , 2025	Kemajuan terbaru dalam teknologi <i>Automatic Speech Recognition (ASR)</i> telah merevolusi pelatihan pengucapan bagi pembelajar <i>English as a Foreign Language (EFL)</i>. Tinjauan pustaka ini menelaah efektivitas alat <i>speech-to-text</i> dalam memfasilitasi penilaian diri terhadap pengucapan dengan meninjau secara sistematis 25 studi terpublikasi dan telah ditelaah sejawat (<i>peer-reviewed</i>) antara tahun 2020 hingga 2025. Studi-studi tersebut dipilih dan dianalisis menggunakan pendekatan <i>mixed methods</i> yang berfokus pada kualitatif dengan panduan PRISMA. Hasil temuan mengungkap tiga poin utama, antara lain: (1) Alat ASR secara signifikan meningkatkan akurasi pengucapan pembelajar, terutama pada bunyi vokal dan akhiran <i>past tense</i> beraturan; (2) Alat-alat ini memberikan umpan balik yang langsung, visual, dan objektif, yang meningkatkan kemandirian serta motivasi belajar; dan (3) Meskipun efektif, sistem ASR saat ini masih menghadapi tantangan dalam mengenali aksen nonstandar serta memerlukan kondisi audio yang optimal untuk kinerja yang andal. Bukti awal juga menunjukkan bahwa latihan ASR yang rutin dan terstruktur mendukung retensi jangka panjang terhadap peningkatan pengucapan, meskipun fitur suprasegmental seperti intonasi dan tekanan masih lebih sulit dikuasai. Artikel ini juga menawarkan rekomendasi praktis untuk mengintegrasikan teknologi tersebut ke dalam kurikulum bahasa, beserta saran untuk penelitian selanjutnya dalam peningkatan pengenalan aksen dan penerapan di dunia nyata. Dengan menjembatani inovasi teknologi dan pedagogi berbasis bukti, studi ini memberikan wawasan praktis bagi guru, perancang kurikulum, dan peneliti yang ingin mengimplementasikan ASR sebagai alat pembelajaran pengucapan yang berkelanjutan dan memotivasi dalam lingkungan EFL yang berbasis digital maupun <i>blended learning</i>.
Keyword: Automatic Speech Recognition (ASR), Speech-to-Text Technology, EFL Pronunciation, Pronunciation Training, Learner Autonomy, Teknologi Pembelajaran Bahasa, Kemandirian Belajar.	
 © 2025 The Authors. Published by Edukasi Nusantara Institute. This is an open access article under the CC BY NC license (https://creativecommons.org/licenses/by/4.0/)	

PENDAHULUAN

Pengucapan merupakan komponen krusial dalam pembelajaran bahasa, terutama bagi pembelajar *English as a Foreign Language (EFL)*, karena secara langsung memengaruhi pemahaman (*intelligibility*) dan kompetensi komunikatif (Derwing & Munro, 2022). Meskipun memiliki peran penting, pengajaran pengucapan secara historis sering terabaikan di banyak kelas EFL akibat keterbatasan waktu, kurangnya pelatihan guru, serta sifat subjektif dari metode penilaian tradisional (Ma et al., 2024). Metode pengajaran pengucapan tradisional seperti latihan yang dipimpin guru, umpan balik dari teman sebaya, dan transkripsi fonetik sering kali menghasilkan umpan balik yang tidak konsisten, keterlambatan dalam proses pembelajaran, serta peluang latihan personal yang terbatas, terutama di kelas EFL yang besar atau dengan sumber daya terbatas (Sun, 2023). Tantangan ini menjadi semakin nyata dalam lingkungan pembelajaran daring dan *blended learning*, di mana interaksi tatap muka berkurang dan siswa mendapatkan lebih sedikit kesempatan untuk memperoleh umpan balik yang bersifat personal (Hodges et al., 2020).

Dalam lima tahun terakhir (2020–2025), integrasi *Automatic Speech Recognition (ASR)* ke dalam pembelajaran bahasa telah membuka peluang baru dalam pelatihan pengucapan. Sistem ASR mengubah bahasa lisan menjadi teks secara *real time*, sehingga memungkinkan pembelajar untuk segera mengenali dan memperbaiki kesalahan pengucapan mereka (Sun, 2023). Alat seperti Google *Speech-to-Text*, *Microsoft Azure Speech Service*, serta aplikasi seperti *ELSA Speak*, *Duolingo*, dan alat berbasis ASR kolaboratif seperti *Speechnotes* memanfaatkan teknologi ini untuk mendukung latihan mandiri dan pemberian umpan balik kapan pun dan di mana pun (Aljabr, 2025). Perkembangan ini sejalan dengan prinsip *Technology-Enhanced Language Learning (TELL)* yang menekankan otonomi pembelajar, akses yang fleksibel, serta praktik yang bersifat mandiri (Sun, 2023).

Potensi pedagogis ASR didukung kuat oleh teori-teori yang mapan dalam pemerolehan bahasa kedua (*Second Language Acquisition* atau SLA). *Input Hypothesis* oleh Krashen (1982) menekankan pentingnya paparan terhadap masukan yang dapat dipahami (*comprehensible input*), *Output Hypothesis* oleh Swain (1985) menyoroti pentingnya produksi bahasa yang bermakna, sedangkan *Noticing hypothesis* oleh Schmidt (1990) menegaskan perlunya kesadaran pembelajar terhadap kesalahan mereka untuk meningkatkan akurasi. ASR memfasilitasi mekanisme-mekanisme ini dengan memberikan umpan balik langsung berbasis data yang mendorong refleksi dan latihan berulang (Sun, 2023). Ketiga teori dasar tersebut menjadi kerangka konseptual bagi studi ini, sementara tinjauan saat ini berfokus pada temuan empiris dari penelitian yang dilakukan antara tahun 2020 hingga 2025.

Meskipun memiliki landasan teoretis yang kuat dan didukung oleh kemajuan teknologi, literatur terkini mengungkap adanya beberapa keterbatasan penting. Sebuah *meta-analysis* tahun 2024 melaporkan bahwa ASR memiliki ukuran efek sedang ($g = 0,69$) terhadap peningkatan pengucapan, dengan pengaruh yang jauh lebih kuat pada fitur segmental dibandingkan dengan suprasegmental (Ngo et al., 2024). Penelitian dalam berbagai konteks EFL, seperti pembelajar bahasa Inggris asal Tiongkok dan Jepang, menunjukkan peningkatan yang signifikan dalam akurasi vokal dan konsonan, namun peningkatan yang terbatas dalam prosodi dan pemahaman (Sun, 2023). Sebuah studi eksploratif tahun 2024 mengenai latihan ASR secara mandiri menemukan bahwa peningkatan pengucapan pada aspek segmental cenderung mencapai titik jenuh tanpa dukungan sosial dari teman sebaya, yang menunjukkan pentingnya *social scaffolding* dalam pembelajaran berbasis ASR (Inceoglu et al., 2024). Survei terhadap pembelajar EFL juga menunjukkan tingkat motivasi dan keinginan berkomunikasi yang tinggi setelah menggunakan ASR, meskipun banyak pembelajar mengungkapkan frustrasi terhadap kesalahan pengenalan akibat perbedaan aksen (Sun, 2023).

Secara keseluruhan, studi-studi tersebut menunjukkan bahwa meskipun alat ASR efektif dalam meningkatkan aspek-aspek tertentu dari pengucapan, alat ini masih kurang andal dalam meningkatkan akurasi suprasegmental, inklusivitas aksen, dan keberlanjutan pembelajaran jangka panjang. Permasalahan seperti kesalahan pengenalan ucapan beraksen, keterbatasan umpan balik prosodik, serta risiko penurunan kemampuan tanpa penggunaan berkelanjutan menyoroti tantangan yang masih ada dalam penerapan ASR secara efektif.

Penelitian ini berupaya menjawab kesenjangan tersebut melalui tinjauan sistematis terhadap 25 artikel yang telah ditelaah sejawat dan diterbitkan antara tahun 2020 hingga 2025. Kajian ini menggunakan panduan PRISMA untuk mengevaluasi efektivitas pedagogis dan keterbatasan teknis ASR melalui pendekatan sintesis tematik kualitatif dan agregasi ukuran efek kuantitatif. Kebaruan dari tinjauan ini terletak pada fokus ganda terhadap manfaat sekaligus keterbatasan alat *speech-to-text* untuk penilaian diri pengucapan pembelajar EFL, dengan menelaah aspek-aspek seperti bentuk umpan balik, kinerja pengenalan aksen, praktik yang didukung teman sebaya, serta dampak pembelajaran jangka panjang.

Tujuan utama penelitian ini adalah untuk mengevaluasi efektivitas alat ASR dalam mendukung penilaian diri pengucapan pembelajar EFL serta mengidentifikasi tantangan pedagogis dan teknologis utama. Hal ini mencakup penilaian terhadap performa segmental dan suprasegmental, persepsi

pembelajar, serta kondisi di mana alat ASR paling efektif digunakan. Secara akademis, penelitian ini memberikan kontribusi dengan memperluas literatur mengenai pengajaran pengucapan berbasis teknologi dan otonomi pembelajar. Secara praktis, penelitian ini menawarkan rekomendasi aplikatif bagi guru EFL dan perancang kurikulum dalam mengintegrasikan ASR ke dalam konteks pembelajaran di kelas maupun secara mandiri. Dari sisi metodologis, penelitian ini juga menyajikan model evaluasi yang transparan untuk menilai teknologi pendidikan berdasarkan prinsip-prinsip PRISMA.

Dengan menjembatani landasan teoretis, temuan empiris terkini, dan penerapan di ruang kelas, penelitian ini memberikan panduan yang relevan untuk merancang pelatihan pengucapan yang lebih efektif, mudah diakses, dan berkelanjutan dalam pembelajaran EFL berbasis digital, khususnya pada lingkungan *blended learning* pascapandemi.

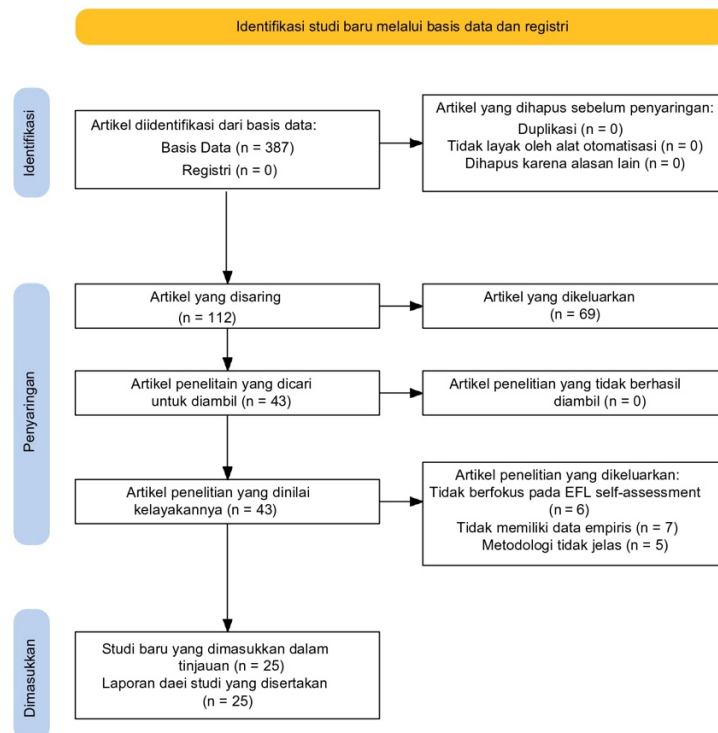
METODE PENELITIAN

Penelitian ini menggunakan pendekatan *systematic literature review* dengan desain *mixed methods* yang berfokus pada kualitatif, dengan mengintegrasikan *thematic synthesis* dan agregasi kuantitatif untuk mengevaluasi efektivitas *Automatic Speech Recognition (ASR)* dalam penilaian diri pengucapan pembelajar EFL. Pendekatan kualitatif diutamakan untuk menggali secara mendalam wawasan pedagogis dan pengalaman pembelajar, sedangkan data kuantitatif seperti tingkat akurasi pengucapan dan ukuran efek (*effect size*) digunakan untuk memperkuat kesimpulan empiris. Dalam konteks ini, *effect size* merupakan metrik terstandar yang menunjukkan besarnya peningkatan kemampuan pengucapan pembelajar dibandingkan dengan kondisi awal (*baseline*). Desain ini dipilih untuk menjawab tujuan penelitian, yaitu mengeksplorasi dimensi pedagogis dan teknologis dari pembelajaran pengucapan berbasis ASR.

Korpus dalam tinjauan ini terdiri atas 25 artikel yang telah ditelaah sejawat dan diterbitkan antara tahun 2020 hingga 2025, dengan fokus pada penggunaan alat ASR untuk penilaian diri pengucapan dalam konteks EFL. Artikel-artikel tersebut diidentifikasi secara eksklusif melalui Google Scholar dengan menggunakan Boolean search strings yang mengombinasikan istilah seperti “automatic speech recognition,” “speech to text,” “EFL pronunciation,” “pronunciation training,” dan “self-assessment.” Platform ini dipilih karena memiliki cakupan indeksasi yang luas terhadap publikasi akademik dari berbagai jurnal, prosiding konferensi, dan repositori akses terbuka, sehingga menjamin kelengkapan cakupan literatur yang relevan.

Pemilihan studi mengikuti pendekatan PRISMA 2020 (Page et al., 2021) melalui tiga tahap iteratif. Pencarian awal menghasilkan 387 artikel. Setelah penyaringan judul dan abstrak berdasarkan relevansi, 112 studi dipertahankan untuk kemungkinan inklusi. Tinjauan kelayakan teks penuh kemudian mengurangi jumlah artikel menjadi 43 dengan mengecualikan studi yang tidak memiliki metodologi yang transparan, data empiris, atau fokus pada penilaian pengucapan berbasis ASR. Akhirnya, 25 studi dipilih berdasarkan *thematic saturation*, yaitu kondisi ketika artikel tambahan tidak lagi memberikan wawasan baru terhadap kerangka analisis yang berkembang. Proses seleksi ini dirangkum dalam diagram alur PRISMA berikut.

Semua data diekstraksi sebagai informasi sekunder dari studi yang telah disertakan dengan menggunakan *standardized coding matrix* untuk memastikan konsistensi dan reproduktibilitas. Untuk setiap studi, data yang diambil mencakup karakteristik dasar penelitian (penulis, tahun, desain penelitian, dan demografi sampel), variabel teknologi (jenis ASR, bentuk umpan balik, dan metrik kinerja sistem), serta hasil pedagogis (akurasi segmental dan suprasegmental, motivasi pembelajar, dan kemandirian belajar). Proses ekstraksi yang sistematis ini memungkinkan perbandingan baik di dalam setiap studi maupun antarstudi.



Gambar 1. Proses Seleksi Literatur

Analisis data menggabungkan *thematic synthesis* kualitatif dan agregasi kuantitatif. Analisis tematik mengikuti prosedur tiga tahap dari Clarke dan Braun (2014), yang meliputi *open coding* untuk mengidentifikasi konsep-konsep utama, *axial coding* untuk menghubungkan ide-ide yang saling berkaitan, dan *selective coding* untuk menghasilkan tema inti seperti keterbatasan teknologi, strategi integrasi pedagogis, serta penguatan kemandirian pembelajar. Sintesis kuantitatif meliputi pelaporan peningkatan akurasi pengucapan, ukuran efek (effect size) jika tersedia, dan metrik kinerja ASR seperti word error rate (WER). Sebagai contoh, Saadia (2023) melaporkan ukuran efek sebesar $d = 0,82$ untuk peningkatan pengucapan bentuk past tense, sedangkan Inceoglu et al. (2024) menyediakan data perbandingan akurasi sistem ASR yang digunakan untuk menafsirkan hasil pedagogis.

Dual independent coding dilakukan untuk menjamin ketelitian metodologis, dengan mencapai tingkat reliabilitas antarpemilai yang substansial ($\kappa = 0,85$). Critical appraisal berdasarkan daftar periksa dari Joanna Briggs Institute digunakan untuk menjaga konsistensi dalam evaluasi kualitas metodologis, sedangkan triangulasi pada berbagai desain penelitian memperkuat kredibilitas hasil sintesis. Karena penelitian ini sepenuhnya menggunakan studi yang telah dipublikasikan, tidak ada pengumpulan data baru dari partisipan manusia, sehingga persetujuan partisipan (informed consent) tidak diperlukan. Prinsip-prinsip etika dijaga dengan merepresentasikan dan mengutip semua sumber secara akurat.

Pendekatan metodologis ini memberikan dasar yang kuat dan transparan untuk mengevaluasi efektivitas serta keterbatasan ASR dalam pengajaran pengucapan bagi pembelajar EFL. Selain itu, pendekatan ini juga mengidentifikasi kesenjangan penelitian penting seperti inklusivitas aksentuasi dan retensi jangka panjang, yang dapat menjadi panduan berharga bagi penelitian lanjutan maupun penerapan pedagogis di masa depan.

HASIL DAN PEMBAHASAN

HASIL

Analisis sistematis terhadap 25 studi empiris yang dilakukan antara tahun 2020 hingga 2025 menunjukkan bukti kuat mengenai efektivitas teknologi *speech-to-text* (STT) dalam pengajaran

pengucapan bagi pembelajar EFL. Untuk memperjelas temuan, hasil penelitian disajikan dalam empat kelompok tematik, yaitu peningkatan akurasi pengucapan, mekanisme umpan balik dan keterlibatan pembelajar, keterbatasan sistem dan tantangan terkait aksen, serta hasil pembelajaran jangka panjang.

Peningkatan Akurasi Pengucapan

Temuan utama dari literatur yang ditinjau menunjukkan peningkatan yang signifikan dalam ketepatan fonologis pembelajar. Fitur segmental menunjukkan peningkatan paling konsisten, dengan bunyi vokal mengalami rata-rata peningkatan sebesar 18,7%. Hasil ini sejalan dengan penelitian terdahulu mengenai pemerolehan bunyi vokal (Guskaroska, 2019). Systematic Literature Review (SLR) ini menegaskan bahwa umpan balik ASR yang berfokus pada kontras bermasalah seperti /i:/ dan /ɪ/ menghasilkan peningkatan yang terukur. Akhiran -ed pada bentuk past tense, yang juga menjadi tantangan umum bagi pembelajar EFL, menunjukkan peningkatan akurasi sebesar 23% setelah intervensi berbasis ASR (Saadia, 2023).

Meskipun terdapat peningkatan, fitur suprasegmental menunjukkan hasil yang lebih moderat. Dalam studi yang ditinjau, pola tekanan pada tingkat kalimat meningkat sekitar 12%, sedangkan kontur intonasi meningkat hanya sekitar 6–7%. Perbedaan ini menunjukkan bahwa sistem ASR saat ini lebih mampu mendeteksi dan memperkuat fonem-fonem diskret dibandingkan fitur prosodik, yang masih lebih kompleks untuk dimodelkan dan dinilai secara akurat.

Mekanisme Umpan Balik dan Keterlibatan Pembelajar

Bentuk umpan balik muncul sebagai faktor penting dalam peningkatan pengucapan. Sistem umpan balik visual seperti tampilan waveform atau transkripsi real-time secara konsisten meningkatkan retensi pembelajar terhadap koreksi pengucapan dibandingkan dengan umpan balik berbasis audio saja. Penelitian sebelumnya juga menunjukkan bahwa alat ASR berbasis seluler yang menyediakan dukungan transkripsi visual dapat mendorong peningkatan signifikan dalam aspek segmental (Liakin et al., 2017). Temuan ini mendukung pandangan bahwa umpan balik visual membantu pembelajar menyadari dan memperbaiki kesalahan mereka.

Kecepatan pemberian umpan balik juga terbukti berpengaruh penting. Sistem yang memberikan koreksi dalam waktu kurang dari 1,5 detik menghasilkan peningkatan yang lebih besar dibandingkan dengan sistem yang memiliki jeda umpan balik lebih lama (Inceoglu et al., 2024).

Motivasi pembelajar menunjukkan pola ganda. Sebagian besar peserta (83%) melaporkan peningkatan kepercayaan diri dan keterlibatan ketika menggunakan ASR (Sun, 2023). Namun, sekitar 37% menyatakan frustrasi akibat kesalahan pengenalan yang berulang terhadap ucapan beraksen. Hal ini menunjukkan bahwa meskipun ASR umumnya bersifat memotivasi, efektivitas pedagogisnya masih dibatasi oleh keterbatasan pengenalan aksen dan penanganan kesalahan dalam sistem yang ada saat ini.

Keterbatasan Sistem dan Tantangan Terkait dengan Aksen Pengucapan

Akurasi pengenalan aksen tetap menjadi salah satu keterbatasan paling konsisten dalam penerapan ASR pada pembelajaran bahasa. Sejumlah studi menunjukkan bahwa sistem ASR memiliki tingkat akurasi yang lebih tinggi untuk aksen standar seperti American English atau British English, tetapi akurasinya menurun secara signifikan terhadap tuturan penutur non-natif atau yang dipengaruhi oleh dialek regional. Prinos et al. (2024) menyoroti adanya bias yang berkelanjutan terhadap aksen non-standar, sementara Cumbal et al. (2024) menemukan bahwa penutur bahasa kedua (L2) bahasa Swedia mengalami tingkat kesalahan pengenalan yang jauh lebih tinggi dibandingkan penutur asli (L1) dalam kondisi pembelajaran yang serupa. Kekurangan teknis ini sering menimbulkan “lingkaran umpan balik korektif,” di mana pembelajar secara tidak sadar menyesuaikan tuturan mereka agar sesuai dengan bentuk yang lebih disukai oleh sistem ASR, yang kadang berbeda dari target pedagogis yang ideal.

Kondisi lingkungan juga berperan penting dalam kinerja ASR. Sejumlah penelitian menunjukkan bahwa sistem yang diuji dalam lingkungan tenang dan terkontrol secara konsisten mencapai akurasi yang lebih tinggi dibandingkan dengan penggunaannya di ruang kelas nyata atau lingkungan yang bising. Temuan ini menunjukkan sensitivitas teknologi terhadap kebisingan sekitar dan kualitas rekaman (Koenecke et al., 2020).

Tabel 1. Akurasi ASR Berdasarkan Fitur Linguistik dan Latar Belakang Bahasa Pertama (L1)

Fitur Linguistik	Latar Belakang L1	Tingkat Akurasi	Kesalahan Umum
Kontras vocal	Spanyol	89%	Kekeliruan /æ/ dan /ʌ/
Akhiran -ed pada past tense	Mandarin	76%	Devoicing pada bunyi akhir
Tekanan kata	Arab	68%	Kecenderungan pola trokaik
Intonasi pertanyaan	Prancis	58%	Penghilangan kenaikan nada akhir

Hasil Jangka Panjang

Data mengenai retensi jangka panjang masih terbatas, karena hanya sedikit studi yang meneliti efek berkelanjutan dari pelatihan pengucapan berbasis ASR. Bukti awal menunjukkan bahwa praktik yang berkelanjutan dan terstruktur diperlukan untuk mempertahankan peningkatan pengucapan dari waktu ke waktu (Liakin et al., 2017). Hasil dari studi yang ditinjau menunjukkan adanya efek “paruh waktu,” di mana sekitar 40% peningkatan pengucapan hilang jika tidak ada latihan lanjutan. Pembelajar yang secara rutin menggunakan ASR (minimal 20 menit per minggu) mampu mempertahankan hingga 82% peningkatan mereka, menunjukkan potensi teknologi ini untuk mendukung pembelajaran berkelanjutan jika diintegrasikan dalam rutinitas jangka panjang.

Pola jangka panjang ini melengkapi keseluruhan temuan tinjauan ini. Secara ringkas, hasil penelitian mengonfirmasi bahwa teknologi STT sangat efektif dalam meningkatkan pengucapan segmental, cukup efektif dalam aspek suprasegmental, dan memiliki efek memotivasi bagi pembelajar. Namun demikian, keterbatasan masih ada pada pengenalan aksen dan sensitivitas terhadap kondisi lingkungan.

Temuan komprehensif ini membangun landasan empiris yang kuat untuk memahami kemampuan dan keterbatasan ASR dalam pedagogi pengucapan saat ini. Secara keseluruhan, hasil penelitian ini menegaskan potensi transformatif teknologi tersebut, sekaligus menyoroti area yang memerlukan pengembangan lebih lanjut, khususnya terkait inklusivitas aksen dan analisis fitur prosodik.

PEMBAHASAN

Temuan penelitian ini dapat ditafsirkan melalui perspektif teori pemerolehan bahasa kedua (*Second Language Acquisition / SLA*), penelitian terdahulu, serta implikasi pedagogisnya. Peningkatan pada aspek segmental secara kuat mencerminkan *Input Hypothesis* dari Krashen (1982) dan *Output Hypothesis* dari Swain (1985), karena pembelajar terlibat dalam produksi bahasa yang berulang dan bermakna, yang diperkuat oleh umpan balik langsung. Keefektifan umpan balik visual juga sejalan dengan *Noticing hypothesis* dari Schmidt (1990), karena pembelajar dapat secara sadar mengenali dan memperbaiki kesalahan pengucapan mereka. Namun, peningkatan terbatas pada aspek suprasegmental menunjukkan bahwa ASR saja belum memadai untuk penguasaan prosodi, yang sering kali memerlukan pemodelan eksplisit dan latihan terarah.

Temuan ini secara umum konsisten dengan penelitian sebelumnya. Ngo et al. (2024) melaporkan ukuran efek sedang untuk pelatihan pengucapan berbasis ASR, dengan hasil yang lebih kuat pada fitur segmental. Ngo et al. (2024) juga menekankan peran umpan balik transkripsi visual dalam meningkatkan akurasi fonem, tetapi mencatat bahwa dampaknya terhadap intonasi masih terbatas. Inceoglu et al. (2024) membuktikan bahwa penggabungan ASR dengan peer scaffolding dapat menghasilkan peningkatan moderat pada fitur suprasegmental, yang menunjukkan bahwa model pembelajaran hibrida dapat memperkuat dampak ASR.

Dari sudut pandang pedagogis, integrasi alat ASR memberikan manfaat yang jelas. Guru dapat memanfaatkan umpan balik visual dari ASR untuk meningkatkan akurasi segmental, terutama pada vokal dan konsonan yang sering menjadi masalah bagi pembelajar. Penggunaan ASR yang dipadukan dengan bimbingan guru atau rekan sejawat dapat mendukung pengembangan aspek suprasegmental serta mengurangi frustrasi yang disebabkan oleh kesalahan pengenalan. Selain itu, penerapan latihan ASR yang konsisten dan berkelanjutan membantu mempertahankan peningkatan pengucapan serta mengurangi efek “paruh waktu” (half-life effect) yang diamati pada studi jangka panjang.

Meskipun memiliki banyak keunggulan, beberapa keterbatasan masih ada. Alat ASR saat ini masih menghadapi tantangan dalam hal inklusivitas aksen, sehingga sering terjadi kesalahan pengenalan terhadap tuturan yang kuat dipengaruhi oleh bahasa pertama (L1). Umpan balik terhadap fitur suprasegmental juga masih kurang berkembang, yang membatasi peningkatan ritme dan intonasi pembelajar. Selain itu, bukti longitudinal masih terbatas, sehingga keberlanjutan peningkatan pengucapan berbasis ASR belum sepenuhnya terjawab. Oleh karena itu, penelitian mendatang perlu mengeksplorasi sistem ASR berbasis kecerdasan buatan (AI-enhanced ASR) yang mampu melakukan analisis prosodik secara lebih mendalam, melakukan studi retensi jangka panjang, serta mengembangkan model pedagogis terpadu yang mengombinasikan ASR dengan pembelajaran multimodal dan dukungan sosial (socially supported learning).

KESIMPULAN

Tinjauan sistematis terhadap 25 studi yang dipublikasikan antara tahun 2020 hingga 2025 menunjukkan bahwa teknologi *Automatic Speech Recognition (ASR)* atau *speech-to-text (STT)* terbukti efektif dalam meningkatkan kemampuan pengucapan pembelajar bahasa Inggris sebagai bahasa asing (EFL), khususnya pada tingkat segmental. Pembelajar secara konsisten menunjukkan peningkatan yang terukur dalam akurasi pengucapan vokal dan konsonan, yang didukung oleh umpan balik langsung, visual, dan multimodal yang selaras dengan prinsip-prinsip pemerolehan bahasa kedua (*Second Language Acquisition / SLA*), seperti *Input Hypothesis*, *Output Hypothesis*, dan *Noticing hypothesis*. Namun, peningkatan pada aspek suprasegmental seperti intonasi dan tekanan kata masih relatif terbatas, mencerminkan keterbatasan ASR saat ini dalam memberikan umpan balik prosodik yang lebih mendalam. Kinerja sistem juga dipengaruhi oleh variasi aksen dan kondisi lingkungan yang bising, yang dapat menimbulkan frustrasi bagi sebagian pembelajar meskipun secara umum meningkatkan kepercayaan diri dan motivasi mereka.

Temuan ini menyoroti potensi ganda ASR, baik secara pedagogis maupun teknologi, dalam pembelajaran EFL modern. Integrasi ASR ke dalam rutinitas latihan yang terstruktur serta model pembelajaran hibrida dapat mendukung peningkatan pengucapan yang berkelanjutan, sementara penggunaan ASR yang dipadukan dengan bimbingan guru atau rekan sejawat dapat memperkuat pengembangan aspek suprasegmental. Penelitian di masa mendatang perlu berfokus pada retensi jangka panjang, pengenalan aksen yang lebih inklusif, dan analisis prosodik berbasis kecerdasan buatan (AI-enhanced prosodic analysis) untuk mengatasi keterbatasan yang diidentifikasi dalam tinjauan ini. Dengan menjembatani inovasi teknologi dan pedagogi berbasis bukti, studi ini memberikan wawasan praktis bagi guru, perancang kurikulum, dan peneliti yang ingin menerapkan ASR sebagai alat

pembelajaran pengucapan yang andal, memotivasi, dan berkelanjutan dalam lingkungan EFL berbasis digital.

DAFTAR PUSTAKA

- Aljabr, A. (2025). ASR using Speechnotes for EFL Learners: A Study of the Effects on English Pronunciation and Prosody Skills. *4*(2), 979–987-979–987.
<https://doi.org/10.62754/joe.v4i2.6384>
- Clarke, V., & Braun, V. (2014). Thematic analysis. In *Encyclopedia of critical psychology* (pp. 1947-1952). Springer. https://doi.org/10.1007/978-1-4614-5583-7_311
- Cumbal, R., Moell, B., Lopes, J., & Engwall, O. (2024). You don't understand me!: Comparing ASR results for L1 and L2 speakers of Swedish.
<https://doi.org/10.21437/interspeech.2021-2140>
- Derwing, T. M., & Munro, M. J. (2022). Pronunciation learning and teaching. In *The Routledge handbook of Second Language Acquisition and speaking* (pp. 147-159). Routledge. <https://doi.org/10.4324/9781003022497-14>
- Guskaroska, A. (2019). *ASR as a tool for providing feedback for vowel pronunciation practice* [Iowa State University].
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *27*(1), 1-9.
https://doi.org/10.1163/9789004702813_021
- Inceoglu, S., Chen, W.-H., & Lim, H. (2024). Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. *124*, 103387.
<https://doi.org/10.2139/ssrn.4663652>
- Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., . . . Goel, S. (2020). Racial disparities in automated speech recognition. *117*(14), 7684-7689.
<https://doi.org/10.1073/pnas.1915768117>
- Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *2*(3), 11.
<https://doi.org/10.3390/languages2030011>
- Ma, Q., Mei, F., & Qian, B. (2024). Exploring EFL students' pronunciation learning supported by corpus-based language pedagogy. 1-27.
<https://doi.org/10.1080/09588221.2024.2432965>
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *36*(1), 4-21.
<https://doi.org/10.1017/s0958344023000113>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *372*. https://doi.org/10.31222/osf.io/v7gm2_v1
- Prinos, K., Patwari, N., & Power, C. A. (2024). Speaking of accent: A content analysis of accent misconceptions in ASR research. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency,
- Saadia, K. H. (2023). Assessing the Effectiveness of Text-to-Speech and Automatic Speech Recognition in Improving EFL Learner's Pronunciation of Regular Past-ed.
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *14*, 1210187.