

Kode Ganda Tunarungu di Live Caption: Interaksi Bahasa Isyarat–Teks Otomatis dan Perbaikan Kesalahan



Check for updates

Rahman Ari Putra¹, Nabila Putri Handayani², Dimas Arya Saputra³, Intan Maharani Kusuma⁴

^a Program Studi Teknik Informatika, Universitas Gadjah Mada, Indonesia

^b Program Studi Pendidikan Luar Biasa, Universitas Muhammadiyah Jakarta, Indonesia

^c Program Studi Sistem Informasi, Institut Teknologi Nasional Bandung, Indonesia

^d Program Studi Teknologi Rekayasa Perangkat Lunak, Sekolah Tinggi Teknologi Bandung, Indonesia

¹ rizkyariputraa@gmail.com; ² nabilaputrihandayani1@gmail.com; ³ dimasaryasaputra@gmail.com;

⁴ intanmaharani Kusuma2@gmail.com

* Corresponding Author

ABSTRACT

Perkembangan teknologi automatic speech recognition (ASR) telah memungkinkan penggunaan live caption sebagai sarana komunikasi bagi komunitas tunarungu dan hard-of-hearing. Meskipun teknologi ini meningkatkan akses terhadap informasi lisan, caption otomatis masih sering mengandung kesalahan transkripsi yang dapat memengaruhi pemahaman pengguna. Penelitian ini bertujuan untuk menganalisis fenomena kode ganda dalam penggunaan live caption oleh pengguna tunarungu, khususnya dalam interaksi antara bahasa isyarat dan teks otomatis serta strategi pengguna dalam menghadapi kesalahan caption. Penelitian ini menggunakan pendekatan kualitatif dengan desain studi interaksi multimodal. Data dikumpulkan melalui observasi interaksi komunikasi yang menggunakan live caption, perekaman video, serta wawancara mendalam dengan partisipan tunarungu. Analisis data dilakukan melalui identifikasi kesalahan caption, analisis interaksi multimodal, dan analisis tematik terhadap strategi pengguna dalam menafsirkan caption. Hasil penelitian menunjukkan bahwa pengguna tunarungu tidak hanya mengandalkan teks caption sebagai sumber informasi utama, tetapi juga memanfaatkan bahasa isyarat dan konteks visual sebagai bagian dari proses pemahaman komunikasi. Interaksi antara bahasa isyarat dan teks caption membentuk praktik kode ganda yang memungkinkan pengguna menafsirkan atau memperbaiki kesalahan caption selama percakapan berlangsung. Temuan ini menunjukkan bahwa pemahaman komunikasi dalam penggunaan live caption bersifat multimodal dan tidak hanya ditentukan oleh akurasi sistem caption. Penelitian ini memberikan kontribusi konseptual terhadap pengembangan teknologi caption yang lebih adaptif terhadap praktik komunikasi komunitas tunarungu.

Copyright © 2026, The Author(s)

This is an open-access article under the CC-BY-SA license



Article History

Received 2026-02-03

Revised 2026-02-15

Accepted 2026-03-20

Published 2026-03-31

Keywords

live caption, tunarungu, bahasa isyarat, automatic speech recognition, komunikasi multimodal

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) dan pemrosesan bahasa alami (Natural Language Processing/NLP) telah membuka peluang baru dalam pengembangan teknologi aksesibilitas bagi penyandang disabilitas, termasuk komunitas tunarungu dan hard-of-hearing (DHH). Sistem berbasis AI memungkinkan proses pengolahan bahasa secara otomatis, seperti pengenalan ucapan, penerjemahan bahasa, serta konversi ucapan menjadi teks secara real-time. Inovasi ini memberikan kontribusi signifikan terhadap peningkatan akses komunikasi dan pembelajaran yang lebih inklusif bagi individu dengan gangguan pendengaran (Asrifan et al., 2025; ZainEldin et al., 2024). Salah satu teknologi yang berkembang pesat dalam konteks tersebut adalah automatic speech recognition (ASR) yang digunakan untuk menghasilkan live caption atau teks transkripsi percakapan secara langsung.

Live caption telah digunakan secara luas dalam berbagai konteks komunikasi, seperti ruang kelas, rapat kerja, diskusi kelompok, dan interaksi sosial sehari-hari. Teknologi ini

How to cite: Putra, R. A., Handayani, N. P., Saputra, D. A., & Kusuma, I. M. (2026) Kode Ganda Tunarungu di Live Caption: Interaksi Bahasa Isyarat–Teks Otomatis dan Perbaikan Kesalahan. *Journal of Modern Social and Humanities*, 2(2), 77-91. <https://doi.org/10.71094/jmsh.v2i2.363>

memungkinkan ujaran yang dihasilkan oleh pembicara diubah menjadi teks secara real-time sehingga dapat dibaca oleh pengguna tunarungu. Beberapa penelitian menunjukkan bahwa penggunaan caption otomatis dapat meningkatkan akses terhadap informasi bagi mahasiswa tunarungu di lingkungan pendidikan tinggi serta membantu mereka mengikuti diskusi kelas secara lebih efektif (Butler et al., 2019). Selain itu, teknologi caption juga telah berkembang menjadi bagian dari berbagai aplikasi komunikasi aksesibel, seperti sistem transkripsi percakapan pada perangkat seluler dan aplikasi komunikasi berbasis teks (Loizides et al., 2020). Penggunaan teknologi ini bahkan telah diperluas dalam lingkungan pembelajaran digital dan media pendidikan yang mengintegrasikan transkripsi otomatis dengan elemen visual lain, termasuk animasi bahasa isyarat (Widiartha et al., 2024) maupun lingkungan virtual yang dirancang khusus untuk meningkatkan aksesibilitas bagi pengguna tunarungu (Paudyal et al., 2019).

Meskipun teknologi ASR memberikan manfaat yang signifikan bagi akses komunikasi, berbagai penelitian menunjukkan bahwa sistem tersebut masih memiliki keterbatasan dalam menghasilkan transkripsi yang akurat. Caption otomatis sering kali mengandung kesalahan pengenalan kata, terutama dalam situasi percakapan spontan, kondisi akustik yang bising, atau ketika terdapat variasi aksen dan kecepatan berbicara (Romero-Fresco & Fresno, 2023). Kesalahan tersebut dapat menyebabkan distorsi makna atau bahkan kehilangan informasi penting dalam percakapan. Dalam konteks pengguna tunarungu, kesalahan ini memiliki implikasi yang lebih besar karena teks caption sering kali menjadi sumber utama informasi linguistik yang tersedia selama interaksi komunikasi.

Penelitian sebelumnya menunjukkan bahwa kesalahan pada caption otomatis dapat memengaruhi tingkat keterpahaman teks bagi pengguna tunarungu, terutama bagi individu dengan tingkat literasi membaca yang berbeda-beda (Kafle & Huenerfauth, 2016). Studi lain juga menunjukkan bahwa kesalahan transkripsi dapat mengganggu proses pemahaman percakapan, terutama ketika kata-kata kunci dalam kalimat tidak dikenali dengan benar oleh sistem ASR (Van Waes et al., 2009). Oleh karena itu, berbagai pendekatan telah dikembangkan untuk meningkatkan kualitas caption dan meminimalkan dampak kesalahan transkripsi.

Beberapa penelitian berfokus pada pengembangan metode koreksi kesalahan caption secara otomatis menggunakan model bahasa dan teknik pemrosesan sinyal (Cardinal & Boulianne, 2009; Cardinal et al., 2007). Selain itu, pendekatan lain memanfaatkan strategi bantuan manusia untuk memperbaiki kesalahan caption secara real-time. Misalnya, penelitian tentang sistem crowdsourced captioning menunjukkan bahwa integrasi antara kecerdasan manusia dan mesin dapat meningkatkan akurasi caption dalam komunikasi langsung (Lasecki et al., 2013; Lasecki et al., 2014). Pendekatan ini kemudian berkembang menjadi sistem yang lebih kompleks yang mengintegrasikan kontribusi manusia dan algoritma pembelajaran mesin dalam proses pembuatan caption secara real-time (Lasecki et al., 2017).

Selain pendekatan koreksi langsung, beberapa penelitian juga mengembangkan metode untuk membantu pengguna memahami caption yang tidak sempurna. Salah satu pendekatan adalah dengan menampilkan tingkat kepercayaan (confidence level) dari hasil pengenalan kata sehingga pengguna dapat menilai keandalan teks yang ditampilkan (Berke, 2017). Penelitian lain mengembangkan model word importance yang menyoroti kata-kata yang paling penting dalam sebuah kalimat untuk membantu pengguna memahami informasi utama meskipun terdapat kesalahan pada kata lain (Kafle, 2019). Pendekatan berbasis keyword spotting juga digunakan untuk membantu manusia memperbaiki caption dengan lebih cepat selama proses transkripsi berlangsung (Gaur et al., 2015).

Di sisi lain, beberapa penelitian menunjukkan bahwa pendekatan yang sepenuhnya otomatis sering kali masih kurang akurat dibandingkan dengan sistem yang melibatkan intervensi manusia. Misalnya, penelitian tentang sistem respeaking captioning menunjukkan bahwa penggunaan operator manusia untuk mengulang ujaran ke dalam sistem ASR dapat

meningkatkan akurasi caption dibandingkan sistem yang sepenuhnya otomatis (Ruiz-Arroyo et al., 2022). Penelitian lain juga mengembangkan berbagai teknik untuk memperbaiki kesalahan caption secara real-time melalui proses editing langsung oleh manusia selama proses transkripsi berlangsung (Wald, 2006; Wald et al., 2007). Selain itu, penelitian tentang koreksi kesalahan caption juga menyoroti pentingnya desain antarmuka yang memudahkan pengguna dalam melakukan editing atau anotasi secara real-time (Wald & Bain, 2007; Bell, 2007).

Meskipun berbagai penelitian telah berfokus pada peningkatan akurasi caption otomatis, sebagian besar studi tersebut lebih menekankan pada aspek teknis sistem dibandingkan dengan praktik komunikasi pengguna tunarungu dalam memanfaatkan caption tersebut. Penelitian tentang pengalaman pengguna menunjukkan bahwa pemahaman caption tidak hanya bergantung pada akurasi teks, tetapi juga pada strategi interpretasi yang digunakan oleh pengguna selama komunikasi berlangsung (Berke et al., 2017). Selain itu, tingkat literasi membaca yang beragam di kalangan pengguna tunarungu juga memengaruhi cara mereka memproses dan memahami caption otomatis (Berke et al., 2018).

Dalam praktik komunikasi sehari-hari, pengguna tunarungu tidak hanya bergantung pada teks caption, tetapi juga menggunakan berbagai sumber informasi lain seperti bahasa isyarat, gerakan tubuh, ekspresi wajah, dan konteks visual untuk memahami pesan yang disampaikan. Penelitian tentang desain aplikasi komunikasi aksesibel menunjukkan bahwa komunikasi antara individu tunarungu dan individu dengar sering kali melibatkan kombinasi berbagai mode bahasa yang saling melengkapi (Seita et al., 2022). Pendekatan multimodal ini juga tercermin dalam berbagai inovasi teknologi komunikasi, seperti penggunaan tampilan caption transparan dalam interaksi tatap muka (Yamamoto et al., 2021) serta sistem komunikasi berbasis perangkat wearable yang memungkinkan caption ditampilkan langsung dalam bidang pandang pengguna (Kang et al., 2024).

Selain itu, berbagai penelitian juga mulai mengembangkan sistem yang mengintegrasikan caption dengan representasi bahasa isyarat digital atau animasi untuk meningkatkan aksesibilitas komunikasi bagi pengguna tunarungu (Bhagwat et al., 2025; Widiartha et al., 2024). Bahkan, dalam beberapa konteks, komunitas tunarungu juga terlibat langsung dalam proses pembuatan caption melalui kolaborasi kelompok, seperti yang ditunjukkan dalam penelitian tentang captioning berbasis komunitas oleh pengguna tunarungu sendiri (Zhang et al., 2016).

Fenomena tersebut menunjukkan bahwa komunikasi tunarungu pada dasarnya bersifat multimodal, di mana makna dibangun melalui integrasi berbagai sistem representasi bahasa. Dalam konteks penggunaan live caption, teks yang dihasilkan oleh sistem ASR tidak selalu dipahami secara langsung sebagai representasi bahasa yang utuh, tetapi sering kali diinterpretasikan melalui interaksi dengan bahasa isyarat dan konteks komunikasi yang lebih luas. Dengan kata lain, pengguna tunarungu dapat memanfaatkan dua sistem representasi bahasa secara bersamaan, yaitu bahasa isyarat dan teks caption. Fenomena ini dapat dipahami sebagai bentuk kode ganda dalam komunikasi tunarungu, di mana makna dibangun melalui interaksi antara dua sistem bahasa yang berbeda namun saling melengkapi.

Meskipun konsep komunikasi multimodal telah banyak dibahas dalam penelitian teknologi aksesibilitas, masih terdapat keterbatasan penelitian yang secara khusus mengkaji interaksi antara bahasa isyarat dan teks caption dalam praktik komunikasi nyata. Sebagian besar penelitian sebelumnya lebih berfokus pada peningkatan akurasi sistem caption atau pengembangan algoritma koreksi kesalahan. Penelitian yang menelaah bagaimana pengguna tunarungu secara aktif menafsirkan, memperbaiki, atau menegosiasikan makna dari caption yang tidak sempurna masih relatif terbatas. Padahal, pemahaman terhadap praktik komunikasi tersebut penting untuk mengembangkan sistem caption yang lebih adaptif terhadap kebutuhan pengguna.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis fenomena kode ganda tunarungu dalam penggunaan live caption, khususnya dalam konteks interaksi antara bahasa isyarat dan teks otomatis serta strategi pengguna dalam menghadapi kesalahan caption. Secara khusus, penelitian ini berupaya menjawab tiga pertanyaan penelitian utama, yaitu: (1) bagaimana bentuk interaksi antara bahasa isyarat dan teks otomatis dalam penggunaan live caption oleh pengguna tunarungu; (2) bagaimana kesalahan dalam caption otomatis memengaruhi proses pemahaman komunikasi; dan (3) strategi apa yang digunakan oleh pengguna tunarungu untuk memperbaiki atau menafsirkan kesalahan caption selama interaksi berlangsung. Dengan mengkaji fenomena tersebut, penelitian ini diharapkan dapat memberikan kontribusi konseptual terhadap pengembangan teknologi caption yang lebih inklusif serta memperkaya kajian linguistik multimodal dalam konteks komunikasi tunarungu.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan desain studi interaksi multimodal untuk menganalisis bagaimana pengguna tunarungu memanfaatkan live caption dalam komunikasi serta bagaimana mereka mengintegrasikan teks caption dengan bahasa isyarat selama proses interaksi. Pendekatan kualitatif dipilih karena penelitian ini berfokus pada pemahaman mendalam terhadap praktik komunikasi dan pengalaman pengguna dalam menggunakan teknologi caption. Selain itu, komunikasi pengguna tunarungu melibatkan berbagai mode bahasa seperti teks, gestur, ekspresi wajah, dan bahasa isyarat sehingga memerlukan pendekatan analisis yang mampu menangkap dinamika interaksi multimodal (Seita et al., 2022).

Partisipan Penelitian

Partisipan penelitian terdiri dari individu tunarungu atau hard-of-hearing yang memiliki pengalaman menggunakan aplikasi caption otomatis dalam komunikasi sehari-hari. Pemilihan partisipan dilakukan melalui purposive sampling dengan mempertimbangkan beberapa kriteria, yaitu: (1) menggunakan bahasa isyarat sebagai salah satu bentuk komunikasi utama; (2) memiliki pengalaman menggunakan aplikasi live caption atau transkripsi otomatis; serta (3) bersedia berpartisipasi dalam kegiatan observasi dan wawancara penelitian.

Pemilihan partisipan juga mempertimbangkan keragaman tingkat literasi membaca, karena penelitian sebelumnya menunjukkan bahwa tingkat literasi dapat memengaruhi cara pengguna tunarungu memahami caption otomatis (Berke et al., 2018). Dengan mempertimbangkan variasi tersebut, penelitian ini diharapkan dapat memperoleh gambaran yang lebih komprehensif mengenai strategi pemahaman caption di kalangan pengguna tunarungu.

Teknologi Caption yang Digunakan

Penelitian ini menggunakan aplikasi live caption berbasis automatic speech recognition (ASR) yang mampu mengubah ujaran lisan menjadi teks secara real-time. Teknologi ini merupakan salah satu bentuk sistem komunikasi aksesibel yang banyak digunakan oleh komunitas tunarungu dalam berbagai konteks interaksi sosial dan pendidikan (Butler et al., 2019; Loizides et al., 2020).

Meskipun sistem ASR terus mengalami perkembangan, berbagai penelitian menunjukkan bahwa teknologi ini masih menghasilkan kesalahan pengenalan kata yang dapat memengaruhi pemahaman pengguna (Kafle & Huenerfauth, 2016). Oleh karena itu, penggunaan sistem caption otomatis dalam penelitian ini juga memungkinkan analisis terhadap bagaimana pengguna tunarungu menghadapi kesalahan transkripsi selama komunikasi berlangsung.

Teknik Pengumpulan Data

Data penelitian dikumpulkan melalui tiga teknik utama, yaitu observasi interaksi, perekaman video, dan wawancara mendalam.

Observasi Interaksi

Observasi dilakukan untuk mengamati secara langsung interaksi komunikasi antara partisipan tunarungu dan mitra komunikasi yang menggunakan live caption. Fokus observasi meliputi penggunaan bahasa isyarat, respons terhadap teks caption, serta strategi komunikasi yang digunakan ketika terjadi kesalahan caption. Pendekatan observasional ini telah digunakan dalam berbagai penelitian tentang teknologi komunikasi aksesibel untuk memahami praktik interaksi pengguna secara lebih natural (Jain et al., 2018).

Perekaman Video dan Dokumentasi Caption

Seluruh interaksi komunikasi direkam menggunakan perangkat video untuk memungkinkan analisis yang lebih mendalam terhadap aspek multimodal komunikasi. Rekaman tersebut mencakup penggunaan bahasa isyarat oleh partisipan, teks caption yang dihasilkan oleh sistem ASR, serta respons komunikasi yang muncul selama percakapan berlangsung.

Selain itu, teks caption yang dihasilkan oleh sistem juga didokumentasikan untuk dianalisis lebih lanjut. Analisis ini mencakup identifikasi jenis kesalahan caption yang muncul selama interaksi, seperti kesalahan penggantian kata, penghilangan kata, maupun kesalahan struktur kalimat. Klasifikasi kesalahan tersebut merujuk pada penelitian sebelumnya mengenai kesalahan dalam sistem caption berbasis pengenalan suara (Van Waes et al., 2009).

Wawancara Mendalam

Setelah proses observasi selesai, wawancara mendalam dilakukan untuk menggali pengalaman partisipan dalam menggunakan live caption serta persepsi mereka terhadap kesalahan caption yang muncul selama komunikasi. Wawancara juga bertujuan untuk memahami strategi yang digunakan oleh pengguna tunarungu dalam menafsirkan atau memperbaiki caption yang tidak akurat. Pendekatan wawancara ini memungkinkan peneliti memperoleh pemahaman yang lebih komprehensif mengenai pengalaman pengguna dalam menggunakan teknologi caption, sebagaimana dilakukan dalam penelitian sebelumnya tentang pengalaman pengguna tunarungu terhadap caption otomatis (Berke et al., 2017).

Teknik Analisis Data

Analisis data dilakukan melalui beberapa tahap. Pertama, teks caption yang dihasilkan oleh sistem ASR dianalisis untuk mengidentifikasi jenis kesalahan transkripsi yang muncul selama interaksi. Analisis ini dilakukan dengan mengklasifikasikan kesalahan berdasarkan kategori yang telah digunakan dalam penelitian sebelumnya tentang kesalahan ASR (Kafle & Huenerfauth, 2016).

Kedua, rekaman video interaksi dianalisis menggunakan pendekatan analisis interaksi multimodal untuk mengidentifikasi hubungan antara penggunaan bahasa isyarat, teks caption, dan respons komunikasi partisipan. Analisis ini bertujuan untuk memahami bagaimana makna dibangun melalui integrasi berbagai mode bahasa selama komunikasi berlangsung.

Ketiga, data wawancara dianalisis menggunakan teknik analisis tematik untuk mengidentifikasi pola strategi yang digunakan oleh partisipan dalam menghadapi kesalahan caption. Analisis ini juga mempertimbangkan temuan penelitian sebelumnya tentang koreksi kesalahan caption secara real-time dan peran pengguna dalam proses tersebut (Wald, 2006; Wald et al., 2007).

Validitas Penelitian

Validitas penelitian dijaga melalui beberapa strategi, yaitu triangulasi metode, member checking, dan analisis komparatif dengan penelitian sebelumnya. Triangulasi metode dilakukan dengan menggabungkan data observasi, rekaman interaksi, dan wawancara sehingga memungkinkan verifikasi silang terhadap temuan penelitian. Member checking dilakukan dengan meminta partisipan untuk meninjau kembali interpretasi peneliti terhadap data yang diperoleh. Selain itu, hasil penelitian juga dibandingkan dengan temuan penelitian sebelumnya tentang captioning real-time dan teknologi aksesibilitas untuk memastikan konsistensi interpretasi data (Takagi et al., 2015; Simpson et al., 2016; Romriell et al., 2009).

HASIL DAN PEMBAHASAN

Teknologi Caption Otomatis dalam Komunikasi Tunarungu

Teknologi automatic speech recognition (ASR) telah menjadi salah satu komponen utama dalam pengembangan sistem komunikasi aksesibel bagi komunitas tunarungu. Teknologi ini memungkinkan ujaran lisan dikonversi menjadi teks secara otomatis dan ditampilkan secara real-time dalam bentuk caption. Implementasi sistem caption otomatis telah digunakan secara luas dalam berbagai konteks komunikasi, termasuk pembelajaran di perguruan tinggi, percakapan sehari-hari, serta interaksi profesional (Butler et al., 2019; Loizides et al., 2020).

Perkembangan teknologi AI dan NLP turut mempercepat peningkatan kemampuan sistem caption otomatis dalam memahami bahasa alami dan menghasilkan transkripsi yang lebih akurat. Integrasi teknologi ini memungkinkan pengembangan sistem pembelajaran bahasa yang lebih inklusif serta mendukung komunikasi yang lebih efektif bagi individu dengan gangguan pendengaran (Asrifan et al., 2025). Selain itu, penelitian terbaru juga menunjukkan bahwa teknologi AI dan deep learning semakin banyak digunakan dalam sistem komunikasi bagi tunarungu, termasuk dalam pengolahan bahasa, pengenalan gestur, dan sistem penerjemahan bahasa isyarat (ZainEldin et al., 2024).

Namun demikian, meskipun teknologi ASR terus berkembang, sistem caption otomatis masih menghadapi berbagai tantangan dalam menghasilkan transkripsi yang sepenuhnya akurat. Kesalahan pengenalan kata masih sering terjadi, terutama dalam situasi percakapan spontan, lingkungan yang bising, atau ketika terdapat variasi aksen dan kecepatan berbicara (Romero-Fresco & Fresno, 2023).

Kesalahan Caption dan Dampaknya terhadap Pemahaman

Kesalahan dalam caption otomatis dapat memengaruhi pemahaman pengguna terhadap informasi yang disampaikan dalam percakapan. Penelitian menunjukkan bahwa kesalahan pengenalan kata dalam sistem ASR dapat menyebabkan distorsi makna atau hilangnya informasi penting dalam kalimat (Kafle & Huenerfauth, 2016). Kesalahan tersebut dapat berupa penggantian kata, penghilangan kata, maupun kesalahan struktur kalimat yang mengubah makna pesan.

Dalam konteks pengguna tunarungu, caption sering kali menjadi sumber utama informasi linguistik selama komunikasi. Oleh karena itu, kesalahan transkripsi dapat memiliki dampak yang lebih signifikan dibandingkan dengan pengguna dengar yang masih dapat mengandalkan informasi audio sebagai sumber tambahan (Van Waes et al., 2009).

Beberapa penelitian juga menunjukkan bahwa tingkat keterpahaman caption tidak hanya dipengaruhi oleh akurasi teks, tetapi juga oleh faktor lain seperti literasi membaca, konteks percakapan, dan strategi interpretasi pengguna (Berke et al., 2018). Penelitian tentang pengalaman pengguna tunarungu terhadap caption otomatis menunjukkan bahwa pengguna

sering kali harus melakukan interpretasi tambahan untuk memahami caption yang tidak sempurna (Berke et al., 2017).

Pendekatan Perbaikan Kesalahan Caption

Berbagai pendekatan telah dikembangkan untuk meningkatkan kualitas caption otomatis dan meminimalkan dampak kesalahan transkripsi. Salah satu pendekatan yang banyak digunakan adalah pengembangan sistem koreksi otomatis berbasis model bahasa yang mampu mendeteksi dan memperbaiki kesalahan caption secara real-time (Cardinal & Boulianne, 2009; Cardinal et al., 2007).

Selain itu, pendekatan lain menggabungkan peran manusia dalam proses pembuatan caption. Misalnya, sistem crowdsourced captioning memungkinkan beberapa individu untuk secara kolaboratif memperbaiki caption yang dihasilkan oleh sistem ASR sehingga meningkatkan akurasi transkripsi secara real-time (Lasecki et al., 2013; Lasecki et al., 2014). Integrasi antara kecerdasan manusia dan mesin dalam sistem caption ini kemudian dikembangkan lebih lanjut melalui pendekatan hybrid yang menggabungkan algoritma pembelajaran mesin dengan kontribusi manusia (Lasecki et al., 2017).

Pendekatan lain yang juga dikembangkan adalah penggunaan teknik keyword spotting untuk membantu manusia memperbaiki caption dengan lebih cepat selama proses transkripsi berlangsung (Gaur et al., 2015). Selain itu, model word importance juga digunakan untuk menyoroti kata-kata penting dalam kalimat sehingga membantu pengguna memahami informasi utama meskipun terdapat kesalahan pada kata lain (Kafle, 2019).

Penelitian lain juga mengembangkan sistem yang menampilkan tingkat kepercayaan (confidence level) dari hasil pengenalan kata dalam sistem ASR. Informasi ini dapat membantu pengguna dalam menilai tingkat keandalan teks yang ditampilkan dalam caption (Berke, 2017).

Komunikasi Multimodal dalam Komunitas Tunarungu

Komunikasi dalam komunitas tunarungu pada dasarnya bersifat multimodal, yaitu melibatkan berbagai bentuk representasi bahasa seperti bahasa isyarat, teks, gestur, ekspresi wajah, dan konteks visual. Penelitian tentang desain aplikasi komunikasi aksesibel menunjukkan bahwa interaksi antara individu tunarungu dan individu dengar sering kali melibatkan kombinasi berbagai mode komunikasi yang saling melengkapi (Seita et al., 2022).

Beberapa inovasi teknologi juga telah mengembangkan sistem caption yang terintegrasi dengan berbagai bentuk representasi visual untuk mendukung komunikasi tatap muka. Misalnya, teknologi see-through caption memungkinkan teks caption ditampilkan pada layar transparan sehingga pengguna tunarungu dapat membaca caption sambil tetap melihat ekspresi wajah pembicara (Yamamoto et al., 2021).

Selain itu, perkembangan teknologi wearable juga memungkinkan caption ditampilkan langsung dalam bidang pandang pengguna selama percakapan berlangsung, sehingga meningkatkan kenyamanan komunikasi bagi pengguna tunarungu (Kang et al., 2024).

Kode Ganda Tunarungu dalam Interaksi Caption

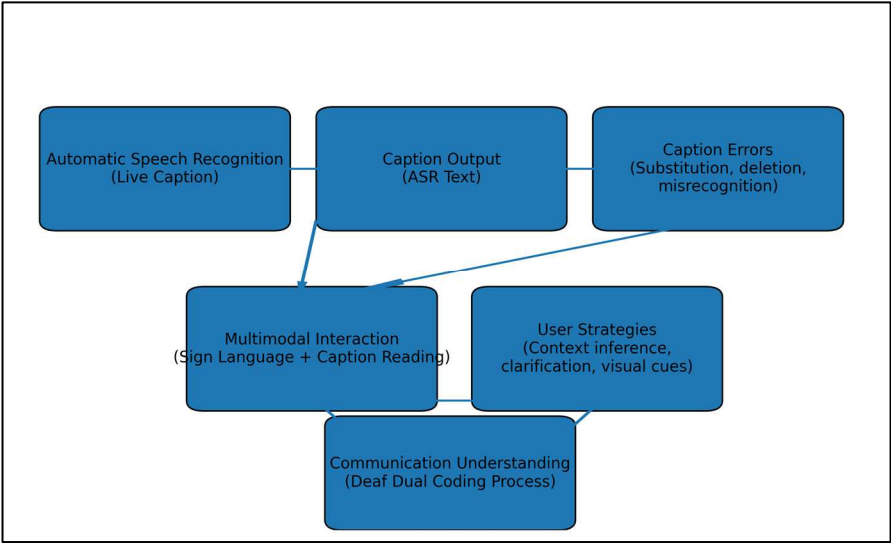
Dalam praktik komunikasi sehari-hari, pengguna tunarungu sering kali memanfaatkan dua sistem bahasa secara bersamaan, yaitu bahasa isyarat dan teks. Interaksi antara kedua sistem bahasa ini dapat dipahami sebagai bentuk kode ganda, di mana makna dibangun melalui integrasi dua representasi linguistik yang berbeda namun saling melengkapi.

Fenomena ini terlihat jelas dalam penggunaan live caption, di mana teks yang dihasilkan oleh sistem ASR tidak selalu dipahami secara langsung sebagai representasi bahasa yang utuh. Sebaliknya, pengguna tunarungu sering kali menginterpretasikan teks caption melalui bantuan bahasa isyarat dan konteks komunikasi yang lebih luas.

Pendekatan komunikasi multimodal ini juga tercermin dalam berbagai inovasi teknologi yang menggabungkan caption dengan representasi bahasa isyarat digital atau animasi (Bhagwat et al., 2025; Widiartha et al., 2024). Bahkan dalam beberapa konteks, komunitas tunarungu juga terlibat secara aktif dalam proses pembuatan caption melalui kolaborasi kelompok pengguna (Zhang et al., 2016).

Dengan demikian, analisis terhadap penggunaan live caption tidak hanya perlu mempertimbangkan aspek teknis sistem caption, tetapi juga praktik komunikasi pengguna tunarungu yang melibatkan integrasi berbagai mode bahasa.

Figure 1. Conceptual Model of Deaf Dual Coding Interaction in Live Caption Communication



Kesalahan Caption sebagai Tantangan Utama dalam Sistem Live Caption

Hasil penelitian menunjukkan bahwa kesalahan dalam caption otomatis merupakan fenomena yang sering muncul dalam penggunaan live caption oleh pengguna tunarungu. Kesalahan tersebut meliputi substitusi kata, penghilangan kata, kesalahan struktur kalimat, serta kesalahan pengenalan istilah atau nama khusus. Temuan ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa sistem automatic speech recognition (ASR) masih memiliki keterbatasan dalam mengenali ujaran secara akurat, terutama dalam situasi percakapan spontan atau lingkungan dengan kualitas audio yang tidak optimal (Romero-Fresco & Fresno, 2023).

Dalam konteks teknologi aksesibilitas, kesalahan caption memiliki implikasi yang lebih signifikan dibandingkan kesalahan transkripsi dalam konteks umum. Bagi pengguna tunarungu, teks caption sering kali menjadi sumber utama informasi linguistik selama percakapan berlangsung. Ketika caption mengandung kesalahan, pengguna tidak memiliki akses langsung terhadap informasi audio untuk mengonfirmasi makna pesan yang disampaikan. Oleh karena itu, kesalahan pengenalan kata dapat menyebabkan distorsi makna yang lebih besar dibandingkan dengan pengguna dengar yang masih dapat mengandalkan informasi suara sebagai sumber tambahan (Kafle & Huenerfauth, 2016).

Penelitian sebelumnya juga menunjukkan bahwa kesalahan caption dapat mengganggu keterpahaman teks terutama ketika kata-kata penting dalam kalimat tidak dikenali dengan benar oleh sistem ASR (Van Waes et al., 2009). Dalam penelitian ini, fenomena serupa juga ditemukan. Beberapa kesalahan caption tidak hanya mengubah bentuk kata, tetapi juga menghasilkan struktur kalimat yang tidak sesuai dengan konteks percakapan. Hal ini

menyebabkan pengguna harus melakukan interpretasi tambahan untuk memahami maksud pembicara.

Temuan ini memperkuat pandangan bahwa meskipun teknologi ASR telah berkembang pesat melalui integrasi kecerdasan buatan dan pemrosesan bahasa alami, sistem caption otomatis masih belum sepenuhnya mampu menggantikan kemampuan manusia dalam memahami bahasa alami yang kompleks (Asrifan et al., 2025; ZainEldin et al., 2024). Oleh karena itu, pemahaman terhadap bagaimana pengguna tunarungu menghadapi kesalahan caption menjadi penting untuk pengembangan sistem caption yang lebih adaptif terhadap kebutuhan pengguna.

Strategi Pengguna Tunarungu dalam Menghadapi Caption Tidak Sempurna

Salah satu temuan penting dalam penelitian ini adalah bahwa pengguna tunarungu tidak hanya bergantung pada teks caption untuk memahami pesan yang disampaikan dalam percakapan. Sebaliknya, mereka mengembangkan berbagai strategi untuk menafsirkan atau memperbaiki kesalahan caption yang muncul selama komunikasi berlangsung.

Strategi pertama yang sering digunakan adalah inferensi kontekstual. Dalam strategi ini, pengguna menafsirkan makna caption berdasarkan konteks percakapan yang sedang berlangsung. Bahkan ketika caption mengandung kesalahan kata, pengguna masih dapat memahami maksud pembicara dengan memanfaatkan konteks topik pembicaraan dan urutan percakapan.

Fenomena ini sejalan dengan penelitian yang menunjukkan bahwa pemahaman caption tidak hanya bergantung pada akurasi teks, tetapi juga pada kemampuan pengguna dalam melakukan interpretasi kontekstual (Berke et al., 2018). Pengguna dengan pengalaman yang lebih luas dalam menggunakan caption otomatis cenderung lebih mampu mengidentifikasi pola kesalahan sistem ASR dan menyesuaikan strategi pemahaman mereka.

Strategi kedua adalah klarifikasi melalui bahasa isyarat atau komunikasi visual. Dalam beberapa situasi, ketika caption tidak dapat dipahami, pengguna tunarungu akan meminta klarifikasi kepada mitra komunikasi melalui bahasa isyarat atau gestur. Hal ini menunjukkan bahwa caption tidak selalu berfungsi sebagai sistem komunikasi yang berdiri sendiri, tetapi menjadi bagian dari sistem komunikasi yang lebih luas yang melibatkan berbagai mode bahasa.

Temuan ini konsisten dengan penelitian tentang pengalaman pengguna tunarungu terhadap caption otomatis yang menunjukkan bahwa pengguna sering kali harus melakukan negosiasi makna selama komunikasi berlangsung (Berke et al., 2017). Dalam penelitian tersebut, caption dipahami sebagai alat bantu komunikasi yang memerlukan interpretasi aktif dari pengguna, bukan sebagai representasi bahasa yang sepenuhnya akurat.

Strategi lain yang ditemukan dalam penelitian ini adalah penggunaan ekspresi wajah dan gestur pembicara sebagai petunjuk tambahan untuk memahami makna caption. Dalam interaksi tatap muka, pengguna tunarungu sering kali menggabungkan informasi dari caption dengan informasi visual dari pembicara untuk membangun pemahaman yang lebih lengkap terhadap pesan yang disampaikan.

Temuan ini mendukung penelitian sebelumnya yang menunjukkan bahwa komunikasi dalam komunitas tunarungu bersifat multimodal dan melibatkan integrasi berbagai bentuk representasi bahasa (Seita et al., 2022).

Kode Ganda dalam Interaksi Bahasa Isyarat dan Caption

Salah satu kontribusi utama penelitian ini adalah identifikasi fenomena kode ganda tunarungu dalam penggunaan live caption. Kode ganda dalam konteks ini merujuk pada penggunaan dua sistem representasi bahasa secara bersamaan, yaitu bahasa isyarat dan teks caption, dalam proses pemahaman komunikasi.

Hasil penelitian menunjukkan bahwa pengguna tunarungu tidak memproses caption secara terpisah dari bahasa isyarat. Sebaliknya, kedua sistem bahasa tersebut saling berinteraksi dalam proses pemaknaan. Ketika caption mengandung kesalahan atau ambigu, pengguna dapat memanfaatkan bahasa isyarat untuk mengklarifikasi makna yang dimaksud oleh pembicara.

Fenomena ini menunjukkan bahwa komunikasi berbasis caption tidak dapat dipahami hanya sebagai interaksi antara pengguna dan sistem teknologi. Sebaliknya, caption merupakan bagian dari praktik komunikasi yang lebih luas yang melibatkan berbagai mode bahasa.

Temuan ini sejalan dengan penelitian yang menunjukkan bahwa komunikasi antara individu tunarungu dan individu dengar sering kali melibatkan kombinasi berbagai mode bahasa yang saling melengkapi (Jain et al., 2018). Dalam konteks tersebut, caption berfungsi sebagai salah satu komponen dalam sistem komunikasi multimodal.

Selain itu, penelitian tentang desain aplikasi komunikasi berbasis caption juga menunjukkan bahwa pengguna tunarungu sering kali menggunakan caption sebagai referensi tambahan untuk memahami percakapan, bukan sebagai satu-satunya sumber informasi (Seita et al., 2022).

Konsep kode ganda yang ditemukan dalam penelitian ini juga berkaitan dengan perkembangan teknologi yang mengintegrasikan caption dengan berbagai bentuk representasi visual. Misalnya, penelitian tentang sistem caption transparan menunjukkan bahwa pengguna tunarungu dapat membaca caption sambil tetap memperhatikan ekspresi wajah pembicara (Yamamoto et al., 2021). Pendekatan ini menunjukkan bahwa caption tidak menggantikan komunikasi visual, tetapi justru memperkuatnya.

Peran Teknologi dalam Mendukung Komunikasi Multimodal

Temuan penelitian ini menunjukkan bahwa teknologi caption tidak hanya berfungsi sebagai alat transkripsi, tetapi juga sebagai komponen dalam ekosistem komunikasi multimodal bagi pengguna tunarungu. Perkembangan teknologi AI dan machine learning telah memungkinkan pengembangan sistem komunikasi yang semakin kompleks dan adaptif terhadap kebutuhan pengguna.

Misalnya, beberapa penelitian telah mengembangkan sistem caption yang terintegrasi dengan animasi bahasa isyarat atau representasi visual lainnya untuk meningkatkan aksesibilitas komunikasi (Bhagwat et al., 2025; Widiartha et al., 2024). Sistem tersebut menunjukkan bahwa integrasi antara teks dan bahasa isyarat dapat meningkatkan efektivitas komunikasi bagi pengguna tunarungu.

Selain itu, teknologi wearable juga mulai digunakan untuk menampilkan caption secara langsung dalam bidang pandang pengguna selama percakapan berlangsung (Kang et al., 2024). Inovasi ini memungkinkan pengguna tunarungu untuk membaca caption tanpa harus mengalihkan perhatian dari pembicara.

Perkembangan teknologi komunikasi aksesibel juga didukung oleh berbagai penelitian tentang penggunaan AI untuk meningkatkan kualitas caption. Misalnya, penggunaan large language models (LLM) telah diusulkan sebagai metode untuk meningkatkan kualitas caption otomatis dengan memperbaiki kesalahan transkripsi secara kontekstual (Fathallah et al., 2024).

Pendekatan Perbaikan Kesalahan Caption

Temuan penelitian ini juga menunjukkan bahwa kesalahan caption tidak selalu menghambat komunikasi secara signifikan karena pengguna tunarungu dapat mengembangkan berbagai strategi untuk menafsirkan kesalahan tersebut. Namun demikian, pengembangan teknologi yang mampu meminimalkan kesalahan caption tetap menjadi aspek penting dalam meningkatkan aksesibilitas komunikasi.

Berbagai penelitian sebelumnya telah mengusulkan berbagai metode untuk memperbaiki kesalahan caption secara real-time. Salah satu pendekatan yang banyak digunakan adalah pengembangan sistem koreksi otomatis berbasis model bahasa yang mampu mendeteksi dan memperbaiki kesalahan transkripsi (Cardinal & Boulianne, 2009; Cardinal et al., 2007).

Pendekatan lain adalah integrasi antara kecerdasan manusia dan mesin dalam sistem caption. Penelitian tentang sistem crowdsourced captioning menunjukkan bahwa kontribusi manusia dapat meningkatkan akurasi caption secara signifikan dibandingkan dengan sistem yang sepenuhnya otomatis (Lasecki et al., 2013; Lasecki et al., 2014). Pendekatan ini kemudian dikembangkan lebih lanjut melalui sistem hybrid yang mengintegrasikan algoritma pembelajaran mesin dengan kontribusi manusia dalam proses pembuatan caption (Lasecki et al., 2017).

Selain itu, penelitian tentang sistem respeaking juga menunjukkan bahwa penggunaan operator manusia untuk mengulang ujaran ke dalam sistem ASR dapat meningkatkan akurasi caption dibandingkan dengan sistem yang sepenuhnya otomatis (Ruiz-Arroyo et al., 2022).

Penelitian lain juga mengembangkan teknik untuk mempercepat proses koreksi caption oleh manusia melalui metode keyword spotting, yang memungkinkan sistem menyoroti kata-kata penting dalam percakapan (Gaur et al., 2015). Selain itu, model word importance juga digunakan untuk membantu pengguna memahami informasi utama dalam caption meskipun terdapat kesalahan pada kata lain (Kafle, 2019).

Peran Pengguna dalam Koreksi Caption

Penelitian ini juga menunjukkan bahwa pengguna tunarungu secara aktif berperan dalam proses koreksi atau interpretasi caption selama komunikasi berlangsung. Dalam beberapa kasus, pengguna tidak hanya membaca caption, tetapi juga melakukan evaluasi terhadap keakuratan teks yang muncul.

Hal ini sejalan dengan penelitian yang menunjukkan bahwa desain antarmuka caption yang baik harus memungkinkan pengguna untuk melakukan koreksi atau anotasi secara real-time (Wald, 2006). Penelitian lain juga menunjukkan bahwa sistem caption yang memungkinkan editing langsung oleh manusia dapat meningkatkan kualitas komunikasi dalam berbagai konteks pendidikan dan profesional (Wald et al., 2007).

Selain itu, penelitian tentang desain tampilan caption juga menunjukkan bahwa personalisasi tampilan dapat meningkatkan keterbacaan dan kemudahan penggunaan sistem caption bagi pengguna tunarungu (Wald & Bain, 2007).

Penelitian lain juga menyoroti pentingnya sistem yang memungkinkan pengguna untuk melihat tingkat kepercayaan dari hasil pengenalan kata dalam caption (Berke, 2017). Informasi ini dapat membantu pengguna dalam menilai apakah teks yang ditampilkan oleh sistem dapat dipercaya atau perlu dikonfirmasi melalui sumber lain.

Implikasi Penelitian

Temuan penelitian ini memiliki beberapa implikasi penting bagi pengembangan teknologi caption dan penelitian tentang komunikasi tunarungu.

Pertama, penelitian ini menunjukkan bahwa pemahaman komunikasi dalam penggunaan live caption tidak hanya ditentukan oleh akurasi sistem caption. Sebaliknya, pemahaman tersebut juga dipengaruhi oleh strategi komunikasi multimodal yang digunakan oleh pengguna tunarungu.

Kedua, penelitian ini menyoroti pentingnya mengembangkan sistem caption yang mendukung integrasi berbagai mode komunikasi, seperti bahasa isyarat, teks, dan visual.

Ketiga, penelitian ini memberikan kontribusi konseptual terhadap kajian linguistik multimodal dengan memperkenalkan konsep kode ganda tunarungu dalam penggunaan live caption.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis fenomena kode ganda dalam penggunaan live caption oleh pengguna tunarungu, khususnya dalam interaksi antara bahasa isyarat dan teks otomatis serta strategi pengguna dalam menghadapi kesalahan caption. Hasil penelitian menunjukkan bahwa penggunaan live caption dalam komunikasi tunarungu tidak dapat dipahami semata-mata sebagai interaksi antara pengguna dan teknologi transkripsi otomatis. Sebaliknya, caption merupakan bagian dari sistem komunikasi yang lebih luas yang melibatkan integrasi berbagai mode bahasa.

Temuan penelitian menunjukkan bahwa kesalahan caption merupakan fenomena yang umum terjadi dalam sistem automatic speech recognition, yang dapat berupa substitusi kata, penghilangan kata, maupun distorsi struktur kalimat. Kesalahan tersebut berpotensi memengaruhi pemahaman pengguna tunarungu karena caption sering kali menjadi sumber utama informasi linguistik selama percakapan berlangsung. Meskipun demikian, penelitian ini menunjukkan bahwa kesalahan caption tidak selalu menyebabkan kegagalan komunikasi karena pengguna tunarungu mampu mengembangkan berbagai strategi interpretasi untuk memahami makna pesan.

Salah satu temuan utama penelitian ini adalah bahwa pengguna tunarungu secara aktif memanfaatkan komunikasi multimodal dalam memahami caption yang tidak sempurna. Strategi yang digunakan meliputi inferensi kontekstual, klarifikasi melalui bahasa isyarat, penggunaan ekspresi wajah dan gestur pembicara, serta negosiasi makna selama percakapan berlangsung. Temuan ini menunjukkan bahwa pemahaman komunikasi dalam penggunaan live caption tidak hanya bergantung pada akurasi sistem caption, tetapi juga pada kemampuan pengguna dalam mengintegrasikan berbagai sumber informasi selama komunikasi.

Penelitian ini juga mengidentifikasi fenomena kode ganda tunarungu, yaitu penggunaan dua sistem representasi bahasa secara bersamaan—bahasa isyarat dan teks caption—dalam proses pemaknaan komunikasi. Dalam praktik komunikasi nyata, teks caption tidak diproses secara terpisah dari bahasa isyarat. Sebaliknya, kedua sistem bahasa tersebut saling melengkapi dalam membangun pemahaman terhadap pesan yang disampaikan. Ketika caption mengandung kesalahan atau ambiguitas, bahasa isyarat dan konteks visual dapat berfungsi sebagai mekanisme klarifikasi yang membantu pengguna memahami maksud pembicara.

Temuan ini menunjukkan bahwa teknologi caption sebaiknya tidak dipandang sebagai pengganti komunikasi visual dalam komunitas tunarungu, tetapi sebagai komponen dalam ekosistem komunikasi multimodal yang melibatkan berbagai bentuk representasi bahasa. Oleh karena itu, pengembangan teknologi caption di masa depan perlu mempertimbangkan praktik komunikasi pengguna tunarungu yang bersifat multimodal dan interaktif.

Secara keseluruhan, penelitian ini menegaskan bahwa efektivitas sistem caption dalam komunikasi tunarungu tidak hanya ditentukan oleh akurasi teknologi ASR, tetapi juga oleh kemampuan sistem tersebut untuk mendukung integrasi berbagai mode komunikasi yang digunakan oleh pengguna.

DAFTAR PUSTAKA

Asrifan, A., Shafa, S., Pratiwi, W. R., et al. (2025). *Bridging gaps: AI and NLP approaches for inclusive language learning*. Advances in Computational Intelligence and Robotics Book Series. <https://doi.org/10.4018/979-8-3693-7260-9.ch006>

- Bell, J.-M. (2007). Enhancing accessibility through correction of speech recognition errors. *ACM SIGACCESS Accessibility and Computing*, (89). <https://doi.org/10.1145/1328567.1328572>
- Berke, L. (2017). Displaying confidence from imperfect automatic speech recognition for captioning. *ACM SIGACCESS Accessibility and Computing*, 118. <https://doi.org/10.1145/3051519.3051522>
- Berke, L., Caulfield, C., & Huenerfauth, M. (2017). Deaf and hard-of-hearing perspectives on imperfect automatic speech recognition for captioning one-on-one meetings. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. <https://doi.org/10.1145/3132525.3132541>
- Berke, L., Kafle, S., & Huenerfauth, M. (2018). Methods for evaluation of imperfect captioning tools by deaf or hard-of-hearing users at different reading literacy levels. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3173574.3173665>
- Bhagwat, S. R., Deshmukh, M. S., Kachhioria, R., et al. (2025). *SmartSignLearn*. IGI Global. <https://doi.org/10.4018/979-8-3693-7560-0.ch025>
- Butler, J., Trager, B., & Behm, B. (2019). Exploration of automatic speech recognition for deaf and hard of hearing students in higher education classes. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. <https://doi.org/10.1145/3308561.3353772>
- Cardinal, P., & Boulianne, G. (2009). Real-time correction of closed captions. In *Proceedings of Interspeech 2009*. <https://doi.org/10.21437/interspeech.2009-443>
- Cardinal, P., Boulianne, G., Comeau, M., & Rouat, J. (2007). Real-time correction of closed captions. In *Proceedings of the Annual Conference of the International Speech Communication Association*. <https://doi.org/10.3115/1557769.1557803>
- Fathallah, N., Bhole, M., & Staab, S. (2024). Empowering the deaf and hard of hearing community: Enhancing video captions using large language models. *arXiv*. <https://doi.org/10.48550/arxiv.2412.00342>
- Gaur, Y., Metze, F., Miao, Y., & Reddy, C. (2015). Using keyword spotting to help humans correct captioning faster. In *Proceedings of Interspeech 2015*. <https://doi.org/10.21437/interspeech.2015-595>
- Jain, D., Franz, R. L., Findlater, L., et al. (2018). Towards accessible conversations in a mobile context for people who are deaf and hard of hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. <https://doi.org/10.1145/3234695.3236362>
- Kafle, S. (2019). Word importance modeling to enhance captions generated by automatic speech recognition for deaf and hard of hearing users.
- Kafle, S., & Huenerfauth, M. (2016). Effect of speech recognition errors on text understandability for people who are deaf or hard of hearing. In *Proceedings of the Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. <https://doi.org/10.21437/slp.2016-4>
- Kang, J. J., Layton, E., Martin, D., et al. (2024). Towards improving real-time head-worn display caption mediated conversations with speaker feedback for hearing conversation partners. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3613905.3650976>
- Lasecki, W. S., Miller, C. D., Kushalnagar, R. S., & Bigham, J. P. (2013). Legion scribe: Real-time captioning by non-experts. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/2461121.2461151>

- Lasecki, W. S., Kushalnagar, R., & Bigham, J. P. (2014). Legion scribe: Real-time captioning by non-experts. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*. <https://doi.org/10.1145/2661334.2661352>
- Lasecki, W. S., Miller, C. D., Naim, I., et al. (2017). Scribe: Deep integration of human and machine intelligence to caption speech in real time. *Communications of the ACM*, 60(11), 93–100. <https://doi.org/10.1145/3068663>
- Loizides, F., Basson, S. H., Kanevsky, D., et al. (2020). Breaking boundaries with live transcribe: Expanding use cases beyond standard captioning scenarios. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. <https://doi.org/10.1145/3373625.3417300>
- Paudyal, P., Banerjee, A., Hu, Y., et al. (2019). DAVEE: A deaf accessible virtual environment for education. In *Proceedings of the 2019 ACM Conference*. <https://doi.org/10.1145/3325480.3326546>
- Romero-Fresco, P., & Fresno, N. (2023). The accuracy of automatic and human live captions in English. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*.
- Romriell, J. N., Brooksby, S. L., Roylance, S. A., et al. (2009). *Methods and systems related to text caption error correction* (Patent).
- Ruiz-Arroyo, A., García-Crespo, Á., & Fuenmayor-Gonzalez, F. (2022). Comparative analysis between a respeaking captioning system and a captioning system without human intervention. *Universal Access in the Information Society*. <https://doi.org/10.1007/s10209-022-00926-3>
- Seita, M., Lee, S., Andrew, S., et al. (2022). Remotely co-designing features for communication applications using automatic captioning with deaf and hearing pairs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3441852.3476551>
- Simpson, M. N., Barrett, J., Bell, P., et al. (2016). Just-in-time prepared captioning for live transmissions. <https://doi.org/10.1049/IBC.2016.0027>
- Takagi, H., Itoh, T., & Shinkawa, K. (2015). Evaluation of real-time captioning by machine recognition with human support. In *Proceedings of the International Conference on Computers Helping People with Special Needs*. <https://doi.org/10.1145/2745555.2746648>
- Van Waes, L., Delbeke, T., Leijten, M., et al. (2009). Live-ondertiteling met spraakherkenning: een case-analyse van reductie en fouten.
- Wald, M. (2006). Captioning for deaf and hard of hearing people by editing automatic speech recognition in real time. In *Lecture Notes in Computer Science*. https://doi.org/10.1007/11788713_100
- Wald, M. (2006). Creating accessible educational multimedia through editing automatic speech recognition captioning in real time. *Interactive Technology and Smart Education*. <https://doi.org/10.1108/17415650680000058>
- Wald, M., Bell, J.-M., & Boulain, P. (2007). Correcting automatic speech recognition captioning errors in real time. *International Journal of Speech Technology*. <https://doi.org/10.1007/s10772-008-9014-4>
- Wald, M., & Bain, K. (2007). Enhancing the usability of real-time speech recognition captioning through personalised displays and real-time multiple speaker editing and annotation. In *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-540-73283-9_50
- Widiartha, K. K., Agustini, K., & Tegeh, I. M. (2024). Real time automated speech recognition transcription and sign language character animation on learning media. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*. <https://doi.org/10.23887/janapati.v13i3.85065>

-
- Yamamoto, K., Suzuki, I., Shitara, A., et al. (2021). See-through captions: Real-time captioning on transparent display for deaf and hard-of-hearing people. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/3441852.3476551>
- ZainEldin, H., Gamel, S. A., Talaat, F. M., et al. (2024). Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-024-10816-0>
- Zhang, J., Shiraishi, Y., Kumai, K., et al. (2016). Real-time captioning of sign language by groups of deaf and hard-of-hearing people. In *Proceedings of the ACM International Conference on Interactive Systems*. <https://doi.org/10.1145/3011141.3011143>