

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Gangguan pendengaran merupakan salah satu disabilitas sensorik paling umum yang memengaruhi kualitas hidup individu, terutama dalam aspek komunikasi sosial dan partisipasi dalam masyarakat. *World Health Organization* (WHO) mencatat bahwa lebih dari 1,5 miliar orang di dunia mengalami gangguan pendengaran dalam berbagai tingkat, dan sekitar 430 juta di antaranya memerlukan layanan rehabilitasi (*World Health Organization*, 2025). Di Indonesia, prevalensi disabilitas pendengaran pada penduduk usia satu tahun ke atas mencapai 0,4%, dan hanya sebagian kecil penyandanganya yang menggunakan alat bantu dengar sebagai sarana komunikasi (Kementerian Kesehatan RI, 2025).

Penyandang tuna rungu dibedakan menjadi dua kategori utama, yaitu tuli kongenital (sejak lahir) dan tuli akuisisi (setelah lahir) (Prevention, 2022). Sebagian besar kasus gangguan pendengaran merupakan tuli akuisisi yang disebabkan oleh faktor usia, infeksi, trauma akustik, maupun penyakit tertentu. Penyandang tuli akuisisi umumnya masih memiliki kemampuan berbicara karena kehilangan pendengaran terjadi setelah proses pemerolehan bahasa. Meskipun demikian, mereka sering mengalami kesulitan dalam memahami tanggapan verbal dari lawan bicara sehingga komunikasi dua arah menjadi kurang efektif dan dapat berdampak pada keterbatasan interaksi sosial serta penurunan kualitas hidup (Suen dkk., 2020).

Berbagai metode telah digunakan untuk membantu komunikasi penyandang tuna rungu, seperti bahasa isyarat, pembacaan gerak bibir, dan alat bantu dengar. Namun, tidak semua penyandang tuna rungu menguasai bahasa isyarat, sementara masyarakat umum juga belum banyak yang memahami bahasa tersebut. Pembacaan gerak bibir memiliki keterbatasan karena dipengaruhi oleh posisi lawan bicara, pencahayaan, dan kemampuan pengguna dalam menginterpretasikan gerakan bibir. Selain itu, alat bantu dengar tidak selalu efektif bagi individu yang mengalami kehilangan pendengaran berat atau total. Oleh karena itu, diperlukan solusi alternatif yang dapat membantu penyandang tuna rungu memahami informasi verbal secara lebih mudah dan praktis (Irdamurni dkk., 2020).

Perkembangan teknologi pengenalan suara (*speech recognition*) memungkinkan suara yang diucapkan seseorang diubah menjadi teks secara otomatis melalui sistem *Speech-to-Text* (STT). Teknologi ini dapat dimanfaatkan sebagai media komunikasi visual bagi penyandang tuna rungu karena informasi percakapan dapat ditampilkan dalam bentuk teks yang mudah dipahami. Penerapan teknologi STT pada berbagai bidang telah menunjukkan hasil yang positif dalam membantu aksesibilitas komunikasi, termasuk pada lingkungan pendidikan dan aktivitas sehari-hari penyandang gangguan pendengaran (Irdamurni dkk., 2020) (Soni, 2025).

Seiring berkembangnya teknologi *wearable*, kacamata pintar (*smart glasses*) menjadi salah satu media yang potensial untuk mengintegrasikan sistem *Speech-to-Text* secara portabel. Berbeda dengan aplikasi pada telepon pintar yang mengharuskan pengguna melihat layar secara terus-menerus, kacamata pintar memungkinkan informasi ditampilkan langsung pada bidang pandang pengguna sehingga komunikasi dapat berlangsung dengan lebih alami. Penelitian dan pengembangan perangkat *wearable* berbasis subtitle percakapan juga menunjukkan bahwa teknologi ini berpotensi meningkatkan aksesibilitas komunikasi bagi penyandang tuna rungu dalam kehidupan sehari-hari (Jain, 2025).

Salah satu tantangan dalam pengembangan perangkat *wearable* berbasis *Speech-to-Text* adalah keterbatasan sumber daya komputasi pada perangkat yang berukuran kecil dan portabel. Model pengenalan suara modern umumnya memerlukan kapasitas pemrosesan yang cukup besar sehingga sulit dijalankan secara optimal pada mikrokontroler berdaya rendah. Pengembangan *Faster-Whisper* memungkinkan proses transkripsi dilakukan dengan lebih cepat dan efisien dibandingkan implementasi *Whisper* konvensional tanpa mengurangi kualitas hasil pengenalan suara secara signifikan (Wang dkk., 2024).

Berdasarkan permasalahan tersebut, penelitian ini merancang dan membangun alat bantu komunikasi visual berupa kacamata pintar berbasis sistem *Speech-to-Text*. Sistem menggunakan mikrofon digital INMP441 untuk menangkap suara percakapan, kemudian data audio dikirim melalui jaringan WiFi menggunakan protokol UDP ke komputer yang menjalankan model *Faster-Whisper* untuk proses transkripsi. Hasil transkripsi selanjutnya dikirim kembali ke

ESP32-S3 dan ditampilkan pada layar OLED yang terpasang pada kacamata sehingga dapat dibaca secara langsung oleh pengguna. ESP32-S3 berfungsi sebagai pengendali utama perangkat sekaligus media komunikasi data, sedangkan sistem catu daya portabel digunakan untuk mendukung mobilitas pengguna. Seluruh komponen dirakit dalam *casing* berbasis 3D *printing* yang ringan dan ergonomis sehingga dapat digunakan secara praktis dalam aktivitas sehari-hari. Dengan demikian, alat yang dikembangkan diharapkan dapat membantu penyandang tuli akuisisi memperoleh informasi percakapan dalam bentuk visual secara lebih mudah, praktis, dan mudah diakses.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

- 1) Bagaimana merancang sistem yang mampu mengenali dan memproses sinyal suara dari mikrofon digital secara akurat sebagai *input* untuk *Speech-to-Text*?
- 2) Bagaimana mengimplementasikan sistem *Speech-to-Text* berbasis ESP32-S3 dan *Faster-Whisper* untuk mengubah suara menjadi teks sebagai media komunikasi bagi penyandang tuna rungu?
- 3) Bagaimana menampilkan hasil transkripsi suara ke layar OLED secara *near real-time* dengan tampilan yang ergonomis dan mudah dibaca oleh pengguna?

1.3 Batasan Masalah

Agar ruang lingkup penelitian tidak terlalu luas, batasan masalah yang diterapkan adalah sebagai berikut:

- 1) Sistem hanya berfungsi untuk menampilkan hasil transkripsi suara dari lawan bicara ke bentuk teks.
- 2) Proses pengenalan suara dilakukan secara *near real-time* menggunakan model *Faster-Whisper* yang dijalankan pada laptop melalui koneksi WiFi dengan dukungan bahasa Indonesia.
- 3) Tampilan teks dibatasi maksimal 6 baris teks menggunakan font 6×8 px pada layar OLED 0.96".
- 4) Komponen utama yang digunakan terdiri dari ESP32-S3, mikrofon digital INMP441, dan layar OLED.

- 5) Sistem tidak mencakup fitur identifikasi pembicara, perintah suara (*voice command*), maupun penyimpanan riwayat percakapan.
- 6) Pengujian dilakukan di lingkungan dalam ruangan (*indoor*) dengan tingkat kebisingan rendah hingga sedang.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

- 1) Merancang sistem akuisisi suara menggunakan mikrofon digital INMP441 dan ESP32-S3 sebagai masukan untuk proses *Speech-to-Text*.
- 2) Mengimplementasikan sistem *Speech-to-Text* berbasis ESP32-S3, komunikasi UDP melalui WiFi, dan *Faster-Whisper* untuk mengubah suara menjadi teks sebagai media komunikasi bagi penyandang tuna rungu.
- 3) Menampilkan hasil transkripsi suara pada layar OLED yang terintegrasi dengan kaca mata pintar secara *near-real-time* sehingga mudah dibaca oleh pengguna.

1.4.2 Manfaat

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan solusi teknologi yang sederhana dan terjangkau untuk membantu penyandang tuna rungu memahami percakapan dalam bentuk teks.
2. Mengurangi hambatan komunikasi bagi penyandang tuna rungu, khususnya pengguna yang masih memiliki kemampuan berbicara namun mengalami kesulitan memahami tanggapan verbal dari lawan bicara.
3. Menghadirkan alat bantu komunikasi visual yang portabel dan mampu menampilkan hasil transkripsi suara secara *near real-time*.
4. Mendorong pemanfaatan teknologi *Speech-to-Text* berbasis *open-source*, seperti *Faster-Whisper*, dalam pengembangan alat bantu bagi penyandang disabilitas.
5. Menjadi referensi bagi penelitian selanjutnya dalam pengembangan perangkat *wearable* berbasis *Speech-to-Text* untuk meningkatkan aksesibilitas dan komunikasi penyandang tuna rungu.