

BAB II. TINJAUAN PUSTAKA

Kemajuan di bidang prosesor, khususnya prosesor daya rendah atau *embedded processor* dan kecerdasan buatan membuat *Human Computer Interaction* (HCI) menjadi ada dimana-mana, bahkan *Automatic speech recognition* juga merupakan bagian dari HCI dan komputasi ubiquitous. *Automatic speech recognition* juga bermanfaat pada aplikasi seperti automated customer service, video game interaktif, dan untuk mengendalikan kendaraan tanpa awak. Meskipun terbilang simpel tidak muluk muluk akan tetapi sudah terpenuhi fungsi utamanya yaitu *speech-to-text*, fungsi yang ditonjolkan disini adalah dapat menggantikan tugas tulis sms cukup dengan mengucapkan kata atau kalimat yang ingin dikirim. Aplikasi ini tidak menyarankan user untuk menggunakan koneksi internet yang lambat [9].

Berbeda dengan penelitian sebelumnya, penelitian yang dilakukan pada tahun 2020 peneliti mempunyai tujuan untuk membandingkan performa speech recognition engine dalam automatic subtitle generation for video dengan bantuan Qt5 peneliti membandingkan tools CMU Sphinx dengan DeepSpeech, peneliti membandingkan word error rate, dan performa tools mulai untuk video playback dan untuk transcription.

Untuk mengoperasikan aplikasi ini hanya perlu menggunakan Qt5 yang merupakan based graphics library C++, dengan menggunakan Qt5 dapat memilih video yang akan diputar, *play, pause, stop*, maupun mengatur volume video[10].

Selain itu ada penelitian yang mempunyai output yang berbeda, pada penelitian yang dilakukan pada tahun 2019 ini menghasilkan bot yang mengesankan, memang jika dilihat dari judulnya sama saja fungsi yang ditonjolkan adalah *speech to text* namun pengaplikasiannya terbilang unik, dalam penelitian ini peneliti menggunakan chatbot interface pada aplikasi chatting yang cukup populer saat ini yaitu telegram, namun sayangnya pada bot ini belum support untuk upload audio atau bahkan video dari local file [11].

Pada tahun yang sama yaitu 2019 juga dilakukan sebuah penelitian yang

menghasilkan aplikasi yang sesuai dengan judul penelitiannya, aplikasi ini dapat menjalankan fungsi *speech to text*, akan tetapi jika dibandingkan dengan penelitian yang lain penelitian ini terbilang terlalu simple, pada penelitian ini hanya dapat live *speech to text* tidak dapat upload file, live yang dimaksud disini adalah hanya dapat merubah suara jadi suara dengan suara yang sekarang yang langsung berbicara tidak bisa menggunakan file suara yang sudah ada sebelumnya [12].

Berdasarkan dari penelitian yang sudah dilakukan sebelumnya, dapat disimpulkan bahwa kegunaan aplikasi pada penelitian sebelumnya bermacam macam namun dapat ditarik garis yang sama yaitu adalah terdapat fungsi atau fitur yang dapat merubah suara menjadi teks. Walau sebenarnya pada penelitian sebelumnya menggunakan beberapa cara yang berbeda untuk mengkonversi suara atau audio menjadi teks, namun outputnya tetap sama yaitu teks.

Perbandingan yang pernah dibangun sebelumnya dengan milik peneliti dapat dilihat pada Tabel 1.1:

Tabel 1. 1 Tabel Perbandingan

Pembanding	Supriyanta and Widodo [9]	Ramani dkk[10]	Shakhovsk a dkk[11]	Nugroho[1 2]	Bagaswanda (2020 *)
Speech-to-text	✓	✓	✓	✓	✓
Video to text	-	✓	-	-	✓
Upload local file	-	-	-	-	✓
Setting output subtitle	-	-	-	-	✓

Setting bahasa	-	-	✓	-	✓
Cross-platform support	-	✓	✓	-	-
User authentication	-	-	✓	-	✓
Platform utama	Mobile	Desktop	Web	Mobile	Web
Bahasa Pemrograman	JAVA	C++	PYTHON	JAVA	PHP, JS

*) Penelitian yang dilakukan