

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi yang semakin meningkat menimbulkan berbagai macam metode ataupun sistem yang memungkinkan komputer mengubah tulisan menjadi suara atau sebaliknya. Sinyal suara manusia mengandung berbagai macam informasi terutama kata atau pesan dalam bentuk ucapan yang akan disampaikan. Dengan semakin berkembangnya teknologi komputer, dimana salah satunya adalah pemanfaatan komputer untuk dapat memproses suara manusia baik pengucapan maupun intonasi yang dikenal dengan istilah prosodi<sup>[10]</sup>.

Pada dasarnya prosodi ini dikelompokkan dalam dua kelompok yaitu pengenalan ucapan atau SR (*Speech Recognition*) dan pensitesa ucapan atau SS (*Speech Synthesis*). Teknologi SR adalah bagian komputer yang dapat memproses dan mengenali ucapan bahasa manusia menjadi suatu teks atau tulisan, sedangkan teknologi SS kebalikannya yaitu bagaimana komputer dapat membangkitkan ucapan teks menjadi suara<sup>[15]</sup>.

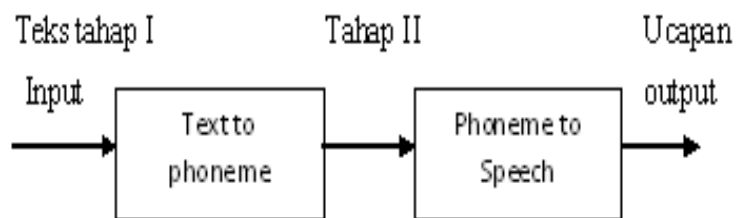
TTS (*Text To speech*) merupakan perancangan dari aplikasi VUI (*Voice User Interface*) yang dikembangkan dari teknologi ASR (*Automatic Speech Recognition*). ASR adalah suatu perkembangan teknologi dan sistem yang memungkinkan komputer untuk menerima masukan berupa kata yang diucapkan. Teknologi ini memungkinkan komputer untuk mengenali suara manusia. VUI merupakan pengembangan dari GUI (*Graphical User Interface*) yang menggunakan visualisasi gambar atau tulisan untuk berkomunikasi antar manusia<sup>[2]</sup>.

*Cool edit pro 2* adalah aplikasi perekaman dan pengolahan data atau audio yang berfungsi sebagai kegiatan *editing* terhadap sebuah gelombang tunggal (*single waveforem*)<sup>[16]</sup> yang menggunakan suara rekaman manusia yang bertujuan untuk penyimpanan data dalam *database* dalam bentuk file wav.

*Microsoft speech API* (SAPI) adalah suatu lapisan perangkat lunak yang digunakan oleh aplikasi pengucapan untuk berkomunikasi dengan *speech*

*recognition (SR) engine* dan *TTS engine*<sup>[12]</sup>. SAPI adalah sistem teks ke ucapan TTS dan pengenalan ucapan *Speech Recognition* sistem TTS mensintesis string teks dan file ke dalam audio ucapan menggunakan suara sintetis pengenalan ucapan untuk mengkonveris audio yang diucapkan manusia ke dalam string teks yang dapat dibaca file.

Pada prinsipnya sistem pensitesa ucapan terdiri dari dua sub sistem, yaitu bagian konverter teks ke fonem (*text to phoneme*) dan bagian konverter fonem ke ucapan (*phoneme to speech*), dimana kedua bagian ini digambarkan oleh gambar 1.1



Gambar 1.1 Prinsip dari Pensintesa Ucapan<sup>[1]</sup>

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas maka yang menjadi rumusan masalahnya adalah:

- Bagaimana memodelkan teks menjadi suara menggunakan *Microsoft Speech API*.
- Bagaimana nilai frekuensi *pitch* dan frekuensi *sampling* menggunakan rekaman *cool edit pro 2* dan menggunakan *Microsoft Speech API*.
- Bagaimana analisis perbandingan nilai frekuensi *pitch* dan frekuensi *sampling* menggunakan rekaman *cool edit pro 2* dan *Microsoft Speech API*.

## 1.3 Batasan Masalah

Pada Tugas Akhir ini masalah akan dibatasi pada:

- Data masukan berupa suara yang direkam dan disimpan dalam *database*.

- b. Tugas Akhir ini mengukur parameter frekuensi *pitch* dan frekuensi *sampling*.
- c. Teknik perekaman suara dalam bentuk file wav agar dapat diolah oleh komputer.
- d. Tugas Akhir ini tidak membahas proses pengenalan suara atau pengenalan pola dan proses sintesis suara.
- e. Tugas Akhir ini tidak membahas algoritma yang digunakan *Microsoft speech API*.
- f. Tugas Akhir ini tidak membahas struktur dan jenis *database* yang digunakan pada matlab R2010a.
- g. Tugas Akhir ini menggunakan *Microsoft speech API* yang *free*.

#### 1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah:

- a. Untuk memodelkan teks menjadi suara menggunakan *Microsoft Speech API*
- b. Mengetahui frekuensi *pitch* dan frekuensi *sampling* menggunakan rekaman *cool edit pro 2*
- c. Mengetahui frekuensi *pitch* dan frekuensi *sampling* menggunakan rekaman *cool edit pro 2* dan *Microsoft Speech API*
- d. Melakukan analisis perbandingan frekuensi *pitch* dan frekuensi *sampling* dari rekaman *cool edit pro 2* dan *Microsoft speech API* (SAPI).

#### 1.5 Posisi Penelitian (*State of the Art*)

Posisi penelitian pada tugas akhir ini ditunjukkan pada gambar 1.2

PERANCANGAN APLIKASI VOICE USER  
INTERFACE DENGAN MENGGUNAKAN  
MICROSOFT SPEECH API

Oleh : Budi Supriyono  
Jurusan Teknik Elektro Fakultas Teknik  
Universitas Diponegoro Semarang  
2011

Pengujian yang dilakukan adalah penggunaan memori oleh aplikasi, pengaruh kondisi lingkungan terhadap jalannya aplikasi, pengaruh kualitas *microphone* pada aplikasi dan tanggapan responden terhadap aplikasi ini.



Gambar 1.2 Posisi Penelitian (*State of the Art*)

## 1.6 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini adalah:

### **BAB I PENDAHULUAN**

Bab ini terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan, metodologi penelitian, serta sistematika penulisan tugas akhir.

### **BAB II DASAR TEORI**

Bab ini membahas mengenai teori-teori yang berkaitan dengan penyelesaian Tugas Akhir, yaitu pengolahan aplikasi VUI, proses pengenalan ucapan, teknologi-teknologi yang mendasari aplikasi *Microsoft speech API* dan *cool edit*.

### **BAB III METODOLOGI PENELITIAN**

Bab ini berisi tentang metodologi penelitian yang digunakan dalam menyusun tugas akhir.

### **BAB IV PERANCANGAN DAN SIMULASI SISTEM**

Bab ini membahas mengenai perancangan sistem, meliputi perancangan sistem, perekaman suara, perancangan *cool edit* dan *Microsoft speech API*.

### **BAB V HASIL SIMULASI DAN ANALISIS**

Bab ini menyajikan hasil simulasi dengan menggunakan Matlab dan analisis hasil perbandingan frekuensi *pitch* dan frekuensi *sampling* dengan menggunakan rekaman *cool edit* dan *Microsoft speech API*.

### **BAB VI PENUTUP**

Bab ini yang berisi kesimpulan dan saran untuk pengembangan Tugas Akhir lebih lanjut.