PEMANFAATAN TEKNOLOGI TEXT-TO-SPEECH SEBAGAI MEDIA PEMBELAJARAN PADA LABORATORIUM BAHASA INGGRIS

Kukuh Yudhistiro, S.Kom.,M.Kom
Fakultas Teknologi Informasi, Universitas Merdeka Malang
kukuh.yudhistiro@unmer.ac.id

Abstract: *English learning laboratory is a very important facility to support the effectiveness and efficiency of the learning process teaching English in schools or other institutions pedidikan. Presence technology speech synthesizer (synthesizer speech), which is also known by the term system Text-To-Speech (TTS), can perform the conversion of the text produced computer in the form of pronunciation (audio), where the pronunciation is generated adjustable speed, intonation (prosody), and the audio output format to be stored in the form of an audio file. This makes saving and keefesienan in the use of tools and without the use of instructional media were quickly worn media such as CDs and devices.*

Keywords: *text-to-speech, speech synthesizer, laboratorium Bahasa Inggris*

Pendahuluan

Laboratorium Bahasa Inggris merupakan fasilitas yang sangat penting untuk menunjang keefektifan dan keefisienan proses belajar mengajar mata pelajaran Bahasa Inggris di sekolah atau lembaga pedidikan lain. Laboratorium ini dilengkapi fasilitas dan perangkat yang memudahkan proses mengajar bagi guru dan belajar bagi murid. Sebagai contoh, laboratorium Bahasa Inggris saat ini memungkinkan seorang guru untuk mengajar listening dan reading dengan memutar materi conversation atau materi pembelajaran lain yang berbentuk kaset, CD, dan VCD melalui sebuah CD player, komputer atau tape recorder yang akan didengarkan oleh murid melalui headset yang sudah dipasang di setiap bangku.

Alamat Korespondensi
Kukuh Yudhistiro
Email: kukuh.yudhistiro@gmail.com

Pesatnya perkembangan materi dan metode pembelajaran Bahasa Inggris, sekolah diperhadapkan dengan kebutuhan untuk memperlengkapi bahan pengajaran seperti CD atau kaset pelajaran Bahasa Inggris yang harganya relatif mahal dan berisiko mengalami kerusakan seiring waktu pemakaian. Masalah penting lain yang terjadi adalah guru kurang fleksibel dalam membuat sendiri materi pembelajaran dalam laboratorium karena keterbatasan sumber dana untuk mengoleksi CD/kaset dan hanya bergantung pada bahan yang ada pada CD/kaset yang sudah tersedia di pasaran saja. Dalam hal ini guru tidak maksimal dalam memberikan materi yang sesuai dengan kebutuhan dan tahap perkembangan siswa.

Kehadiran teknologi *speech synthesizer* (pensintesa ucapan), yang dikenal pula dengan istilah sistem *Text-To-Speech* (TTS), dapat melakukan konversi dari teks yang dihasilkan komputer ke dalam bentuk pengucapan (audio), dimana pengucapan yang dihasilkan dapat diatur kecepatannya, intonasi (prosodi), serta format audio outputnya untuk disimpan dalam bentuk

file audio. Teknologi TTS diharapkan semakin mengefektifkan proses belajar dan mengajar serta melengkapi media pembelajaran mata pelajaran Bahasa Inggris khususnya pada laboratorium Bahasa Inggris di sekolah.

Berdasarkan latar belakang tersebut maka permasalahan dalam penelitian dapat dirumuskan sebagai berikut:

1. Bagaimana membuat sistem belajar dan mengajar yang efektif dan efisien pada laboratorium Bahasa Inggris sekolah?
2. Bagaimana cara menciptakan media pengajaran Bahasa Inggris pada laboratorium Bahasa Inggris dengan program aplikasi (software pensintesa suara) berbasis teknologi *Text-To-Speech*?

Manfaat

Berikut ini adalah manfaat-manfaat yang akan didapat dari penelitian software pensintesa ucapan pada laboratorium Bahasa Inggris:

1. Bagi guru dan sekolah

a. Pada saat mengajar di laboratorium Bahasa Inggris, guru dapat menggunakan program aplikasi media pembelajaran yang dapat mengkonversi input berupa tulisan berbahasa Inggris pada komputer ke dalam bentuk suara atau pengucapan secara otomatis. Sistem seperti ini akan membuat mengajar lebih efektif dan efisien.
b. Guru dapat memasukkan tulisan atau artikel berbahasa Inggris yang disesuaikan dengan metode mengajarnya atau kebutuhan siswa ke dalam software untuk diucapkan secara otomatis oleh komputer.
c. Guru dapat mengatur kecepatan (rate) pengucapan teks dan menghentikan (pause) proses pengucapan dimana saja pada suatu kalimat, sehingga akan sangat menolong bagi siswa yang masih kesulitan dalam hal listening maupun speaking untuk belajar secara perlahan.
d. Guru dapat menyimpan hasil pengucapan teks tersebut ke dalam format audio, sehingga dapat disimpan dalam sebuah compact disc atau media penyimpanan lain untuk diputar kembali dan bersifat portable.
e. Dengan adanya program pensintesa suara ini, guru tidak lagi kesulitan mencari materi pengajaran, tidak lagi hanya bergantung pada CD/kaset pembelajaran Bahasa Inggris yang ada, sehingga guru dapat mencari materi berupa teks atau membuat materi sendiri.
f. Dapat menghemat pengeluaran untuk pembelian CD/kaset pembelajaran Bahasa Inggris yang relatif mahal.
g. Karena sistem pembelajaran ini tidak memakai hardware pemutar piringan CD atau kaset, maka hal ini sangat efisien dalam mengurangi resiko aus atau kerusakan pada perangkat-perangkat keras tersebut.

2. Bagi siswa:

a. Siswa terlatih dalam hal listening karena suara yang dihasilkan speech synthesizer ini merupakan rekaman ribuan diphone dari orang berbahasa Inggris murni yang terdapat dalam sebuah database diphone.
b. Siswa sendiri dapat membuat latihan listening sendiri sesuai kebutuhannya.

Hasil dan Pembahasan

Teknologi Text-to-speech dapat mengkonversi sejumlah kalimat dalam bentuk teks berbahasa tertentu menjadi sebuah pengucapan dalam bahasa yang sama. Text-to-speech berbahasa Inggris akan mengkonversi teks berbahasa Inggris ke dalam pengucapan berbahasa Inggris pula.

Text-To-Speech pada dasarnya berfungsi sebagai mesin yang mengkonversi teks ke dalam PCM (Pulse Code Modulation) audio digital. Unsur-unsur dari sistem Text-to-speech adalah:

1. Text normalization
2. Homograph disambiguation
3. Word pronounciation
4. Prosody
5. Concatenate wave segments

Berikut adalah penjelasan tiap unsur dari sistem *text-to-speech*:

1. *Text Normalization*

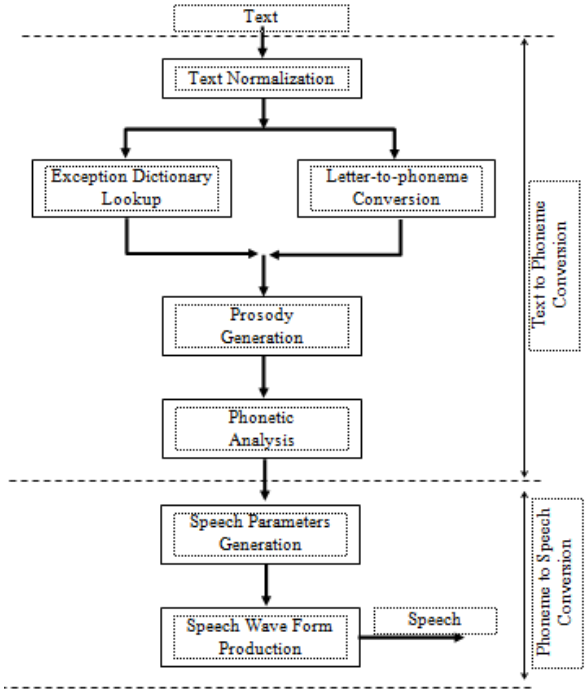
Komponen ini mengkonversi sejumlah masukan (input) berupa teks ke dalam sebuah rangkaian pengucapan kata. Sederhananya, normalisasi teks bekerja mengkonversi suatu kalimat seperti "John rode home." ke sebuah rangkaian kata-kata "John", "rode", "home". Proses normalisasi akan lebih rumit pada kalimat "John rode home at 23.5 mph", dimana "23.5 mph" akan dikonversi menjadi "twenty three point five miles per hour". Pada saat inilah kita dapat mengetahui bagaimana normalisasi teks bekerja.

Pertama, normalisasi teks mengisolasi kata-kata yang ada dalam sebuah teks. Proses ini cukup sederhana, yaitu menganalisa urutan karakter alfabet, dengan mengecek adanya karakter atau simbol seperti apostrophe dan tanda penghubung pada teks. Normalisasi teks mencari karakter bertipe angka, waktu, tanggal, dan simbol-simbol yang lain, kemudian dianalisa dan dikonversi ke dalam bentuk kata atau kalimat. Sebagai contoh: "$54.32" akan dikonversi menjadi "fifty four dollars and thirty two cents". Penkonversian tersebut sangat bergantung pada bahasa yang dipakai.

Berikutnya adalah pengkoversian teks yang berupa singkatan atau abbreviation. Contohnya adalah "in" yang merupakan singkatan dari "inches", "cm" singkatan dari "centimeter" dan "St." singkatan dari "street" dan "saint". Komponen normalisasi teks akan menggunakan sebuah database singkatan (database of abbreviations) dan mencari kepanjangan dari singkatan yang dianalisa. Normalisasi teks juga berlaku pada bentuk teks pada alamat internet. Pada teks alamat internet www.microsoft.com selalu diucapkan "w w w dot microsoft dot com".

Untuk tanda baca, normalisasi teks dapat mendeteksi apakah tanda baca tersebut menyebabkan sebuah kalimat boleh diucapkan atau berhenti. Sebagai contoh tanda baca titik (".."). Pada setiap akhir kalimat, tanda baca titik tidak akan diucapkan, namun pada alamat internet tanda baca titik selalu diucapkan "dot".

Tahapan-tahapan utama konversi dari teks menjadi ucapan dapat dinyatakan dengan diagram seperti terlihat pada gambar berikut



Urutan proses konversi dari teks ke ucapan (Pelton, 1992)

Tahap berikutnya adalah melakukan konversi dari teks yang sudah secara

lengkap merepresentasikan kalimat yang ingin diucapkan menjadi kode-kode fonem. Konversi teks menjadi fonem biasanya dilakukan dengan dua cara. Sebagian proses konversi dapat dilakukan dengan aturan konversi yang sederhana dan berlaku umum untuk berbagai kondisi. Sebagian proses lainnya bersifat kondisional, tergantung dari huruf-huruf atau fonem-fonem tetangganya, bahkan terdapat bentuk-bentuk translasi yang tidak dapat ditemukan keteraturannya.

Konversi yang teratur dapat diimplementasikan dengan tabel konversi yang berisi pasangan antara urutan huruf dan urutan fonem, bahkan mungkin hanya berisi satu huruf dan satu fonem. Aturan yang lebih sulit biasanya diimplementasikan dengan tabel konversi yang akan diterapkan jika kondisi rangkaian huruf tetangga kiri dan kanannya terpenuhi. Contoh bentuk aturan konversi huruf ke fonem yang memenuhi teknik tersebut adalah sebagai berikut.

Left-context [*letter-set*] *right-context = phoneme string*

Huruf tertentu yang ditunjuk dalam posisi [letter-set] akan dikonversikan menjadi suatu fonem dalam "phoneme string" jika left-context dan right context terpenuhi.

Bahasa Inggris termasuk bahasa yang mempunyai keteraturan yang rendah untuk proses konversi teks ke fonem. Suatu TTS bahasa Inggris biasanya dilengkapi dengan suatu basis data yang berisi ribuan kata serta konversi padanan urutan fonemnya. Bahasa Indonesia termasuk bahasa yang jelas aturan konversinya. Sebagian besar kata dalam Bahasa Indonesia dapat dikonversikan menjadi fonem dengan aturan yang jelas dan sederhana, walaupun tetap ada kondisi-kondisi yang tidak dapat ditemukan keteraturannya. Sebagai contoh, simbol huruf e dapat diucapkan sebagai e pepet atau e taling, artinya harus dikonversikan menjadi fonem yang berbeda untuk kondisi yang berbeda. Dalam blok diagram di atas, kondisi yang masih dapat ditangani oleh aturan diimplementasikan dengan blok Letter to Phoneme Conversion. Konversi yang tidak teratur ditangani oleh bagian Exception Dictionary Lookup.

Hasil dari tahap tersebut adalah rangkaian fonem yang merepresentasikan bunyi kalimat yang ingin diucapkan. Bagian prosodi generator akan melengkapi setiap unit fonem yang dihasilkan dengan data durasi pengucapannya serta pitch-nya. Data durasi serta pitch diperoleh berdasarkan kombinasi antara tabel atau database serta model prosodi. Secara simbolik, hasil dari bagian ini sudah menghasilkan informasi yang cukup untuk menghasilkan ucapan yang diinginkan.

Satu tahap berikutnya yang masih sering dilakukan adalah Phonetic Analysis. Tahap ini dapat dikatakan sebagai tahap penyempurnaan, yaitu melakukan perbaikan di tingkat bunyi. Sebagai contoh, dalam bahasa Indonesia, fonem /k/ dalam kata bapak tidak pernah diucapkan secara tegas, atau adanya sisipan fonem /y/ dalam pengucapan kata alamiah antara fonem /i/ dan /a/.

2. Homograph disambiguation

Dalam bahasa Inggris maupun bahasa yang lain, terdapat ratusan kata yang memiliki persamaan tulisan, tetapi berbeda pengucapan.

Suatu contoh dalam bahasa Inggris adalah "read", dimana “read” dapat dilafalkan sebagai "reed" atau "red" tergantung pada arti yang dimaksud. Homograph adalah kata yang memiliki persamaan teks tetapi berlainan arti dan pengucapan satu sama lain. Homograph dapat terjadi pada kata-kata, angka dan singkatan. "Ft." mempunyai pengucapan berbeda pada "Ft. Wayne" dan "100 ft". Demikian juga dengan digit angka "1997", yang diucapkan “nineteen ninety seven" jika menyatakan tahun, atau "one thousand nine hundreds and ninety seven" jika menyatakan banyaknya orang atau obyek.

Text-to-speech memiliki bermacam-macam teknik untuk mengatasi kata-kata yang ambigu tersebut. Agar pengucapannya sesuai dengan isi dan maksud dari teks, dilakukan dengan cara mengecek hubungan antar kata dalam kalimat tersebut atau yang disebut konteks. Pada saat konteks kalimat sudah diketahui, cukup mudah untuk memperkirakan pengucapan yang benar.

Cara yang yang paling mudah untuk menghasilkan pengucapan yang benar adalah dengan melihat ke bagian akhir dari kalimat atau dengan melihat pada leksikon.

3. Word Pronounciation

Modul pronounciation (pengucapan kata) menerima teks, dan mengeluarkannya dalam bentuk urutan fonem. Mirip saat kita melihat kamus. Untuk menghasilkan pengucapan kata-kata, pertama-tama sistem text-to-speech mencari dalam sebuah lexicon (kamus pengucapan kata). Jika kata tersebut tidak ada dalam leksikon, maka sistem akan menggantinya dengan pengucapan yang sesuai dengan alfabet yang ada atau disebut dengan istilah letter-to-sound. Leksikon menyimpan kata dan pengucapannya, seperti kata:

Hello → h eh l oe

Sebuah algoritma digunakan untuk membagi sebuah kata dan menggambarkannya sebagai huruf yang menghasilkan suara. Kita dapat secara jelas mengetahui bahwa “h” pada “hello” menghasilkan fonem “h”, “e” menghasilkan fonem “eh”, huruf “l” yang pertama menghasilkan fonem “l”, huruf “l” yang kedua tidak menghasilkan fonem, dan huruf “o” menghasilkan fonem “oe”. Tentu saja pda kata yang lain juga menghasilkan fonem yang berbeda juga, seperti “e” pada “he” menghasilkan fonem “ee”.

4. Prosody

Tanpa prosodi suara yang dihasilkan seperti suara robot, dan prosodi yang buruk akan membuat suara menjadi seperti suara seorang yang sedang mabuk.

Untuk mendapatkan ucapan yang lebih alami, ucapan yang dihasilkan harus memiliki intonasi (prosody). Secara kuantisasi, prosodi adalah perubahan nilai pitch (frekuensi dasar) selama pengucapan kalimat dilakukan atau pitch sebagai fungsi waktu. Pada prakteknya, informasi pembentuk prosodi berupa data-data pitch serta durasi pengucapannya untuk setiap fonem yang dibangkitkan. Nilai-nilai yang dihasilkan diperoleh dari suatu model prosodi. Prosodi bersifat sangat spesifik untuk setiap bahasa, sehingga model yang diperlukan untuk membangkitkan data-data prosodi menjadi sangat spesifik juga untuk suatu bahasa. Beberapa model umum prosodi pernah dikembangkan, tetapi agar digunakan pada suatu bahasa masih perlu banyak penyesuaian yang harus dilakukan. Berikut adalah teknik bagaimana menghasilkan

prosodi: sistem mengidentifikasi awal dan akhir dari sebuah kalimat. Dalam Bahasa Inggris, pitch selalu menurun di akhir kalimat peryataan dan menaik pada akhir kalimat tanya. Besarnya volume dan kecepatan pengucapan meninggi pada saat awal pengucapan dan menurun pada kata terakhir ketika pengucapan sudah berhenti.

5. *Concatenate wave segments*

Bentuk pensintesa digital yang berkembang pada awalnya adalah pensintesa yang dikenal dengan istilah formant synthesizer, bekerja dengan cara mensimulasikan komponen-komponen frekuensi utama pembentuk ucapan yang disebut formant. Salah satu pensintesa ucapan jenis ini yang populer dan banyak digunakan pada berbagai aplikasi adalah cascade-parallel formant synthesizer yang pertama kali diusulkan oleh Dennis Klatt pada tahun 1990. Synthesizer tersebut merupakan pengembangan dari generasi sebelumnya yang juga dirancang oleh Klatt pada tahun 1980. Pensintesa formant tidak dapat menghasilkan suara dengan tingkat kealamian yang tinggi, sehingga perkembangan text-to-speech mengarah pada pencarian alternatif untuk menggunakan pendekatan yang dapat menghasilkan ucapan yang lebih alami. Seiring dengan kecepatan prosesor serta media penyimpanan komputer yang semakin tinggi, pendekatan tersebut mengarah pada sistem yang melakukan penggabungan segmen-segmen ucapan yang direkam sebelumnya. Berdasarkan berbagai pertimbangan teknis dan kualitas yang ingin dicapai, bentuk segmen yang dianggap paling optimum dan banyak digunakan adalah diphone atau dua fonem yang berurutan. Pendekatan dengan cara penyusunan ucapan dari diphone ini disebut diphone concatenation. Tantangan teknis utama pada teknik diphone concatenation adalah mencari algoritma untuk menggabungkan diphone dengan diphone lainnya, serta algoritma untuk memanipulasi diphone, khususnya untuk mengubah durasi serta pitch diphone.

Concatenation diphone dilakukan dengan merekam diphone dari suara asli dari manusia untuk disimpan dalam sebuah database diphone. Dalam bentuk yang sederhana, sistem akan menerima fonem untuk diucapkan, mengambil audio digital dari sebuah database diphone, mengatur pitch, waktu, dan volume, serta mengirimnya ke sound card.

Design Program

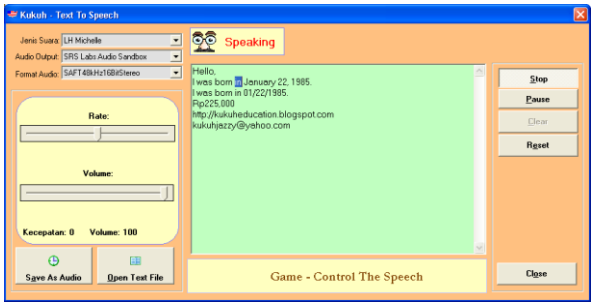
Program aplikasi tersebut dapat digunakan pada komputer dengan spesifikasi:

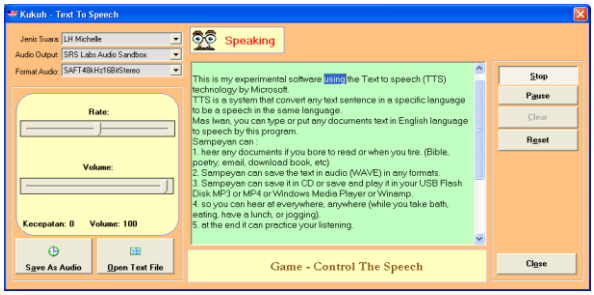
- PC: Microsoft Windows 98, NT4.0, 2000, ME, XP
- Pentium II 200 Mhz atau di atasnya
- 64 MB RAM – minimal 128 untuk XP
- Ruang 100 MB Harddisk
- Mouse dan keyboard
- Output speaker

Berikut adalah tampilan dari aplikasi text-to-speech yang penulis buat:

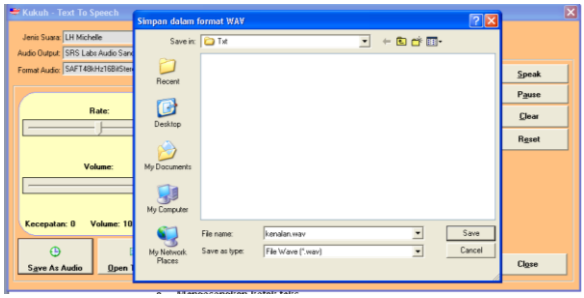
User dapat memasukkan inputan teks berbahasa Inggris baik berbentuk kata maupun paragraph pada kotak teks yang disediakan (hijau). Untuk memulai proses sintesa kalimat user dapat menekan tombol

Start dan setiap kalimat yang sedang dieja/dibunyikan akan diblok. Pensintesa juga mengenali dan membunyikan teks berupa numeric maupun alamat website.





Disamping itu baik guru maupun murid selain dapat mensintesa secara langsung dari aplikasi, mereka dapat menyimpan hasil sintesa tersebut ke bentuk audio, sehingga dapat diperdengarkan kembali dimanapun dan kapanpun dibutuhkan lagi.



Simpulan

1. Siswa terlatih dalam hal listening karena suara yang dihasilkan speech synthesizer ini merupakan rekaman ribuan diphone dari orang berbahasa Inggris murni yang terdapat dalam sebuah database diphone.
2. Siswa sendiri dapat membuat latihan listening sendiri sesuai kebutuhannya.
3. Guru lab dapat membuat dan memodifikasi materi menurut kebutuhan.

Daftar Rujukan

[1] Dutoit, Thierry.1997. "An Introduction to Text-to-Speech Synthesis", Dordrecht: Kluwer Academic Publisher.
[2] http://www.microsoft.com_____
[3] Parsons, Thomas W.1986. "Voice and Speech Processing", New York: McGraw-Hill.
[4] Pelton, Gordon E.1993. "Voice Processing", New York: McGraw-Hill.
[5] Thayer, Rob.1998. "Visual Basic 6 Unleashed", Indiana: Sams.