

APLIKASI TEXT TO SPEECH UNTUK TUNA NETRA MENGUNAKAN METODE DHIPONE CONCATENATION

Ahmad Fahrudi Setiawan, S.Kom, MT

Abstrak - Telah dilakukan penelitian yang berjudul Aplikasi text to speech tuna netra menggunakan metode metode dhipone concatenation yang bertujuan membangun software untuk konversi dari text ke suara yang dapat mengeluarkan suara dari inputan text yang disediakan. Program dapat digunakan untuk membantu orang tuna netra menikmati teknologi dengan cara memasukkan text kedalam program komputer untuk mendapatkan suara..

Program ini dibangun dengan konsep desktop yang dapat digunakan dengan mudah karena hanya menyediakan 1 buah form. Di dalam program ini user dapat mencopy paste text kedalam program dengan cepat dan mendengarkan suara dari hasil text yang telah di copy pastekan.

Secara umum proses dalam sistem TTS terdiri dari *dhipone concatenation* yang berupa modul konversi teks ke *fonem* yang menghasilkan transkripsi fonetik beserta informasi intonasi dan ritme (dikenal dengan prosodi) dan *Digital Signal Processing* (DSP) yang berupa modul konversi *fonem* ke ucapan, yang mengubah informasi fonetis yang diterimanya menjadi sinyal ucapan.

Kata kunci : *Text to speech, tuna netra, dhipone concatenation.*

I : PENDAHULUAN

1.1 Latar Belakang

Membaca adalah jendela dunia, dengan membaca kita memasuki pintu ilmu pengetahuan. Para ilmuan membaca setiap hari untuk bisa mengerti, memahami dan menjadi pintar. Tetapi membaca hanya bisa dilakukan oleh manusia yang bisa melihat, sedangkan tuna netra tidak dapat membaca karena tidak dapat melihat.

Tuna netra menggunakan pendengaran untuk mengetahui suatu hal karena mereka tidak bisa membaca. Dengan ketidak mampuan mereka mengakses buku dan text visual maka mereka tidak memiliki akses dalam ilmu pengetahuan. Dari keterbatasan tuna netra tersebut ada celah yang dapat dilakukan untuk menolong para tuna netra agar bisa seolah-olah membaca dengan membuat *Audio-Learning*.

Audio-learning merupakan pembelajaran dengan memanfaatkan teknologi elektronik yang berupa suara, yaitu hanya memilih *digital book* apa yang ingin di dengarkan dengan hanya mengklik *digital book* tersebut maka kompjuter langsung membacakannya. Teknik yang digunakan adalah *diphone concatenation* dengan prinsip yang terdiri dari dua sub sistem, yaitu: Bagian konverter text ke fonem (*text to phoneme*), serta Bagian konverter fonem ke ucapan (*phoneme to speech*).

Bagian Konverter Teks ke Fonem berfungsi untuk mengubah kalimat masukan dalam suatu bahasa tertentu yang berbentuk teks menjadi rangkaian kode-kode bunyi yang biasanya direpresentasikan dengan kode fonem, durasi serta *pitch*-nya. Bagian ini bersifat sangat *language dependant*. Untuk suatu bahasa baru, bagian ini harus dikembangkan secara lengkap khusus untuk bahasa tersebut.

Bagian Konverter Fonem ke Ucapan akan menerima masukan berupa kode-kode fonem serta *pitch* dan durasi yang dihasilkan oleh bagian sebelumnya. Berdasarkan kode-kode tersebut, bagian Konverter Fonem ke Ucapan akan menghasilkan bunyi atau sinyal ucapan yang sesuai dengan kalimat yang ingin diucapkan. Ada beberapa alternatif teknik yang

dapat digunakan untuk implementasi bagian ini. Dua teknik yang banyak digunakan adalah *formant synthesizer*, serta *diphone concatenation*.

Formant synthesizer bekerja berdasarkan suatu model matematis yang akan melakukan komputasi untuk menghasilkan sinyal ucapan yang diinginkan. Synthesizer jenis ini telah lama digunakan pada berbagai aplikasi. Walaupun dapat menghasilkan ucapan dengan tingkat kemudahan interpretasi yang baik, synthesizer ini tidak dapat menghasilkan ucapan dengan tingkat kealamian yang tinggi.

Berdasarkan penelitian-penelitian yang telah dilakukan, maka dapat diasumsikan bahwa Synthesizer yang menggunakan teknik *diphone concatenation* dapat menghasilkan bunyi ucapan dengan tingkat kealamian (*naturalness*) yang tinggi. Maka dalam tugas akhir ini akan dirancang dan dibuat aplikasi audio-learning dengan menggunakan metode diphone concatenation.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka yang menjadi rumusan masalah yaitu bagaimana membuat *text to speech* menggunakan metode *diphone concatenation*?

1.3 Batasan Masalah

Metode yang digunakan dalam *text to speech* ini adalah Metode *diphone concatenation*.

II. TINJAUAN PUSTAKA

2.1 Pengertian Text To Speech

Sistem konversi *text-to-speech* (TTS) merupakan suatu sistem yang mampu memproduksi sinyal ucapan secara otomatis melalui transkripsi grafem-ke-fonem untuk kalimat yang diucapkan. Perbedaan sistem TTS dengan *talking machine* biasa adalah keotomatisannya dalam mengucapkan kata-kata baru. Oleh karena itu TTS memungkinkan untuk diimplementasikan pada bidang aplikasi yang beragam seperti aplikasi sms bicara, buku digital dan pembaca email otomatis.

Pada prinsipnya, sistem TTS (*Text to Speech*) adalah suatu sistem yang dapat melakukan konversi dari *text* menjadi ucapan. Saat ini terdapat berbagai sistem yang dapat melakukan fungsi seperti itu, misalnya sistem IVR (*interactive voice response*) yang banyak digunakan untuk layanan informasi otomatis melalui telpon (*call center*). Sistem IVR biasanya menggunakan rekaman kata atau kalimat yang direkam secara utuh. Untuk mempertegas definisi TTS dalam Tugas akhir ini, berikut ini akan ditinjau beberapa definisi TTS dari beberapa literatur.

Dutoit dalam buku "*An Introduction to Text-to-Speech Synthesis*" [Dut97] mendefinisikan *Text-to-Speech* sebagai "*production of speech by machines, by way of the automatic phonetization of the sentences to utter*". Pelton dalam buku "*Voice Processing*" [Pel93] menyatakan bahwa "*The task of Text to Speech system is to convert plain text into speech*". Dalam bagian lainnya, Pelton menyatakan pula "*A very attractive advantage of text to speech is that any text can be read, vocabulary is not restricted to utterances that have been decided upon beforehand*".

Speech FAQ, suatu situs Internet yang merangkum pendapat dari berbagai universitas, lembaga penelitian dan industri di bidang aplikasi ucapan, menyatakan bahwa "*Speech synthesis programs convert written input to spoken output by automatically generating synthetic speech. Speech synthesis is often referred to a Text-to-Speech conversion (TTS)*". Situs ini terletak di URL <http://www.speech.cs.cmu.edu/comp.speech/index.html> dengan

penanggung jawab Andrew Hunt dari *Speech Applications Group, Sun Micro systems Laboratories*.

Berdasarkan sumber-sumber tersebut, suatu sistem TTS dapat didefinisikan sebagai “Sistem yang dapat mengubah suatu teks menjadi ucapan secara otomatis dengan cara *fonetisasi* (penyusunan fonem-fonem untuk membentuk ucapan)”. Sebuah sistem TTS dapat mengucapkan kata apapun, dan kosa katanya tidak terbatas. Berdasarkan definisi tersebut, maka suatu sistem IVR tidak dapat dikategorikan sebagai sistem TTS, karena hanya dapat mengucapkan kata-kata atau kalimat dengan jumlah atau kombinasi yang sangat terbatas, tidak dapat mengucapkan semua kata atau kalimat. Selain itu, IVR tidak melakukan *fonetisasi*.

Secara umum proses dalam sistem TTS terdiri dari *Natural Language Processing* (NLP) yang berupa modul konversi teks ke *fonem* yang menghasilkan transkripsi fonetik beserta informasi intonasi dan ritme (dikenal dengan prosodi) dan *Digital Signal Processing* (DSP) yang berupa modul konversi *fonem* ke ucapan, yang mengubah informasi fonetis yang diterimanya menjadi sinyal ucapan.

2.2 Pentingnya Bahasa

Pentingnya bahasa bagi manusia kiranya tidak perlu diragukan lagi. Hal itu tidak saja dapat dibuktikan dengan menunjuk pemakaian bahasa dalam kehidupan sehari-hari, tetapi dapat dibuktikan dengan menunjuk banyaknya perhatian para ilmuwan dan praktisi terhadap bahasa. Bahasa sebagai objek ilmu tidak dimonopoli oleh para ahli bahasa. Para ilmuwan dalam bidang lain pun menjadikan bahasa sebagai objek studi karena mereka memerlukan bahasa sekurang-kurangnya sebagai alat untuk mengkomunikasikan berbagai hal.

Politisi mempelajari bahasa agar dapat menemukan ciri kata atau kalimat dan gaya bahasa yang dapat menyentuh hati nurani orang-orang di sekitarnya sehingga dapat mempengaruhi mereka. Para ahli ilmu jiwa (psikologi dan psikiater) mempelajari bahasa agar dapat menemukan kata-kata yang dapat berperan dalam penyembuhan pasiennya. Dengan anggapan bahwa *speech therapy* mempunyai daya sugestif terhadap hilangnya penyakit, dokter-dokter pun perlu mempelajari bahasa. Untuk mendekatkan diri dengan masyarakat di tempat bertugas, para pamong, para paneliti, para penyuluh sering mempelajari bahasa daerah setempat untuk memudahkan mereka berinteraksi sosial demi kelancaran tugasnya. Bahasa juga dipelajari oleh wartawan, seniman, usahawan, dan oleh orang-orang dari beraneka profesi untuk mengungkapkan pikiran, pandangan, perasaan, dan berbagai maksud lainnya.

Selaku makhluk sosial yang memerlukan orang lain sebagai mitra berkomunikasi, manusia memang memakai cara berkomunikasi, yaitu secara verbal dan nonverbal. Berkomunikasi secara verbal dilakukan dengan menggunakan alat/media bahasa (lisan dan tulisan), sedangkan berkomunikasi secara nonverbal dilakukan dengan menggunakan media selain bahasa. Alat komunikasi nonverbal yang wujudnya berupa aneka simbol, isyarat, kode, dan bunyi. Misalnya tanda lalu lintas, morse, lambaian tangan, sirine, kentongan, atau terompet. Barulah bermakna setelah “diterjemahkan” ke dalam bahasa manusia. Hal itu menunjukkan bahwa merupakan alat komunikasi yang terpenting bagi manusia.

2.3 Fungsi Bahasa

Dalam literatur bahasa, para ahli umumnya merumuskan fungsi bahasa bagi setiap orang ada empat. Yaitu: sebagai alat berkomunikasi, sebagai alat mengekspresikan diri,

sebagai alat berintegrasi dan beradaptasi sosial, sebagai alat kontrol sosial, (keraf 1994: 3-6). Kalau kita cermati, sebenarnya ada satu lagi fungsi bahasa yang selama ini kurang disadari oleh sebagian anggota masyarakat, yaitu sebagai alat untuk berfikir. Seperti kita ketahui, ilmu tentang cara berfikir adalah logika. Dalam proses berfikir, bahasa selalu hadir bersama logika untuk merumuskan konsep, proposisi, dan simpulan. Segala kegiatan yang menyangkut perhitungan atau kalkulasi, pembahasan atau analisis, bahkan berangan-angan atau berkhayal, hanya dimungkinkan berlangsung melalui proses berfikir disertai alatnya yang tidak lain adalah bahasa.

Sejalan dengan uraian diatas dapat diformulasikan bahwa semakin tinggi kemampuan berbahasa seseorang, makin makin tinggi pula kemampuan berfikirnya; makin teratur bahasa seseorang, makin teratur pula cara berfikirnya. Dengan berpegang pada formula itulah di dalam prakata buku ini penulis berani mengatakan seseorang tidak mungkin menjadi intelektual tanpa menguasai bahasa. Seorang intelektual pasti berfikir, dan proses berfikir pasti memerlukan bahasa.

2.4 Bahasa

Bahasa adalah sistem lambang bunyi ujaran yang digunakan untuk berkomunikasi oleh masyarakat pemakainya. Bahasa yang baik berkembang berdasarkan suatu sistem, yaitu seperangkat aturan yang dipatuhi oleh pemakainya. Bahasa sendiri berfungsi sebagai sarana komunikasi serta sebagai sarana integrasi dan adaptasi.

Berikut ini adalah pengertian dan definisi bahasa menurut para ahli: Menurut Bill Adams “Bahasa adalah sebuah sistem pengembangan psikologi individu dalam sebuah konteks inter-subjektif”. Wittgenstein “Bahasa merupakan bentuk pemikiran yang dapat dipahami, berhubungan dengan realitas, dan memiliki bentuk dan struktur yang logis”. Ferdinand De Saussure “Bahasa adalah ciri pembeda yang paling menonjol karena dengan bahasa setiap kelompok sosial merasa dirinya sebagai kesatuan yang berbeda dari kelompok yang lain”.

Plato “Bahasa pada dasarnya adalah pernyataan pikiran seseorang dengan perantaraan onomata (nama benda atau sesuatu) dan rhemata (ucapan) yang merupakan cermin dari ide seseorang dalam arus udara lewat mulut”. Bloch & Trager “Bahasa adalah sebuah sistem simbol yang bersifat manasuka dan dengan sistem itu suatu kelompok sosial bekerja sama”. Carrol “Bahasa adalah sebuah sistem berstruktur mengenai bunyi dan urutan bunyi bahasa yang sifatnya manasuka, yang digunakan, atau yang dapat digunakan dalam komunikasi antar individu oleh sekelompok manusia dan yang secara agak tuntas memberi nama kepada benda-benda, peristiwa-peristiwa, dan proses-proses dalam lingkungan hidup manusia”.

Sudaryono “Bahasa adalah sarana komunikasi yang efektif walaupun tidak sempurna sehingga ketidaksempurnaan bahasa sebagai sarana komunikasi menjadi salah satu sumber terjadinya kesalahpahaman”. Saussure “Bahasa adalah objek dari semiologi”. Mc. Carthy “Bahasa adalah praktik yang paling tepat untuk mengembangkan kemampuan berpikir”. William a. Haviland “Bahasa adalah suatu sistem bunyi yang jika digabungkan menurut aturan tertentu menimbulkan arti yang dapat ditangkap oleh semua orang yang berbicara dalam bahasa itu”.

Bila dilihat dari beberapa definisi dan pengertian mengenai bahasa menurut beberapa ahli diatas, kita bisa melihat bahwa terdapat perbedaan definisi tentang bahasa dimana definisi dari setiap ahli tergantung dengan apa yang ingin ditekankan oleh setiap tersebut. Namun meskipun terdapat perbedaan, nampaknya disepakati bersama bahwa bahasa

adalah alat komunikasi. Dan sebagai alat komunikasi , bahasa mempunyai fungsi-fungsi dan ragam-ragam tertentu.

2.5 Sejarah Bahasa Indonesia

Bahasa Indonesia adalah bahasa resmi Republik Indonesia dan bahasa persatuan bangsa Indonesia. Bahasa Indonesia diresmikan penggunaannya setelah Proklamasi Kemerdekaan Indonesia, tepatnya sehari sesudahnya, bersamaan dengan mulai berlakunya konstitusi. Di Timor Leste, bahasa Indonesia berstatus sebagai bahasa kerja.

Dari sudut pandang linguistik, bahasa Indonesia adalah salah satu dari banyak ragam bahasa Melayu. Dasar yang dipakai adalah bahasa Melayu Riau (wilayah Kepulauan Riau sekarang) dari abad ke-19. Dalam perkembangannya ia mengalami perubahan akibat penggunaannya sebagai bahasa kerja di lingkungan administrasi kolonial dan berbagai proses pembakuan sejak awal abad ke-20. Penamaan "Bahasa Indonesia" diawali sejak dicanangkannya Sumpah Pemuda, 28 Oktober 1928, untuk menghindari kesan "imperialisme bahasa" apabila nama bahasa Melayu tetap digunakan. Proses ini menyebabkan berbedanya Bahasa Indonesia saat ini dari varian bahasa Melayu yang digunakan di Riau maupun Semenanjung Malaya. Hingga saat ini, Bahasa Indonesia merupakan bahasa yang hidup, yang terus menghasilkan kata-kata baru, baik melalui penciptaan maupun penyerapan dari bahasa daerah dan bahasa asing.

Meskipun dipahami dan dituturkan oleh lebih dari 90% warga Indonesia, Bahasa Indonesia bukanlah bahasa ibu bagi kebanyakan penuturnya. Sebagian besar warga Indonesia menggunakan salah satu dari 748 bahasa yang ada di Indonesia sebagai bahasa ibu. Penutur Bahasa Indonesia kerap kali menggunakan versi sehari-hari (kolokial) dan/atau mencampuradukkan dengan dialek Melayu lainnya atau bahasa ibunya. Meskipun demikian, Bahasa Indonesia digunakan sangat luas di perguruan-perguruan, di media massa, sastra, perangkat lunak, surat-menyurat resmi, dan berbagai forum publik lainnya, sehingga dapatlah dikatakan bahwa Bahasa Indonesia digunakan oleh semua warga Indonesia.

Fonologi dan tata bahasa Bahasa Indonesia dianggap relatif mudah. Dasar-dasar yang penting untuk komunikasi dasar dapat dipelajari hanya dalam kurun waktu beberapa minggu.

2.6 Abjad

Abjad yang digunakan dalam bahasa Indonesia terdiri atas 52 huruf, yaitu 26 huruf besar (A-Z) dan 26 huruf kecil (a-z).

2.7 Natural Language Processing (NLP)

Modul NLP dapat diimplementasikan dengan beberapa solusi, yang sering diklasifikasikan sebagai *dictionary-based* dan *rule-based*. Solusi *dictionary-based* diimplementasikan dengan menyimpan sebanyak mungkin informasi *fonologi* ke dalam kamus. Dalam metoda ini transkripsi dilakukan dengan cara metoda *lookup database* leksikal yang telah disusun. Sedangkan sistem transkripsi *rule-based*, menggantikan penyimpanan informasi fonologi dalam kamus dengan membuat set aturan *letter-to-sound* (atau *grafem-ke-fonem*).

2.8 Synthesizer

Tahap pemrosesan terakhir dari sistem TTS adalah sintesa sinyal ucapan. Secara umum terdapat tiga metoda dasar untuk sintesa sinyal ucapan. Sintesis *articulatory*, yang

berusaha memodelkan sistem produksi sinyal ucapan manusia dengan pendekatan fisik mekanis secara langsung, *formant synthesizer* yang memodelkan frekuensi pole suatu sinyal ucapan atau fungsi transfer yang berbasis vocal track atau model source-filter, *synthesizer* perangkaian (*concatenation*), yang menggunakan panjang bagian yang berbeda dari suatu perekaman sinyal ucapan alami. Namun demikian dua teknik yang sering digunakan adalah *formant synthesizer* dan *diphone concatenation*.

Formant synthesizer didasarkan pada pemodelan *source-filter*, merupakan pensintesis dengan pendekatan deskripsi *acousticphonetic*. Model dasar akustik berupa model sumber atau filter. Filter dibuat dengan beberapa set formant, yang menggambarkan artikulasi dalam suatu ucapan suara. Sumber menggambarkan *phonation*. Model ini memodelkan aliran suara (glottal flow) atau sinyal eksitasi noise. Baik sumber dan filter dikontrol oleh suatu set aturan fonetis (biasanya ratusan). Walaupun dapat menghasilkan ucapan dengan tingkat kemudahan interpretasi yang baik, namun *synthesizer* ini tidak dapat menghasilkan ucapan dengan tingkat kealamian yang tinggi.

Synthesizer perangkaian bekerja dengan merangkai unit ucapan yang telah tersimpan sebelumnya (database unit ucapan). Merangkai rekaman bagian ucapan alami merupakan cara termudah untuk membuat sinyal ucapan yang jelas dan alami. Tetapi, pesintesis perangkaian hanya terbatas pada seorang pengucap, dan satu suara, dan biasanya membutuhkan kapasitas memori yang lebih banyak daripada metoda yang lain. Kata (*word*) mungkin menjadi unit ucapan yang paling alami untuk suatu tulisan teks tetapi dengan keterbatasan perbendaharaan kata. Perangkaian kata relatif lebih mudah untuk dibuat dan pengaruh koartikulasi antar kata dapat diatasi. Dengan banyaknya kata yang harus tercakup, maka penggunaan kata sebagai unit ucapan kurang cocok untuk sistem TTS. Jumlah suku kata yang berbeda pada suatu bahasa biasanya lebih kecil dari pada kata, tetapi banyaknya unit dalam database masih terlalu banyak untuk sistem TTS.

Modul sintesis adalah bagian akhir dari sistem TTS yang bertugas mengkonversi informasi fonetik dan prosodi yang ada menjadi sinyal ucapan.

2.9 Fonem

Fonem dapat juga digunakan sebagai unit ucapan pada *database*. Umumnya unit dasar ini berjumlah antara 40-50, yang jauh lebih kecil jika dibandingkan dengan unit yang lain. Penggunaan *fonem* memberikan fleksibilitas yang tinggi dengan sistem *rule-based*. Tetapi, beberapa *fonem* yang tidak memiliki target posisi yang tetap, seperti *plosives*, akan sulit untuk disintesis. Artikulasi juga harus dirumuskan sebagai aturan. Beberapa masalah pada sintesis perangkaian dibandingkan dengan metoda yang lain, yaitu:

1. Terjadi distorsi akibat ketidakberlanjutan pada titik perangkaian, yang dapat dikurangi dengan menggunakan *diphone* atau beberapa metoda lainnya untuk memperhalus sinyal ucapan.
2. Kebutuhan terhadap memori sangat tinggi, khususnya ketika menggunakan unit perangkaian yang panjang, misalnya suku kata dan kata.
3. Pengumpulan data dan penandaan bagian sinyal ucapan membutuhkan waktu yang lama.

Masalah-masalah tersebut dapat diatasi dengan salah satu metode yang ada yaitu *Pitch Synchronous Overlap Add* (PSOLA).

2.10 Konversi dari Text to Speech

Text to Speech (TTS) diartikan sebagai proses pengubahan teks menjadi audio digital dan diucapkan. Pengucapan ini dapat berupa pengiriman audio digital tersebut ke pengeras

suara computer atau menyimpan hasil pengubahan tersebut untuk diputar nanti. Tugas sistem TTS secara umum dapat dibagi dalam 2 bagian besar, analisa teks dan sintesa ucapan. Analisa teks mentransformasi teks masukan menjadi representasi linguistik, selanjutnya bagian sintesis ucapan mentransformasi representasi linguistik tersebut menjadi gelombang sinyal ucapan.

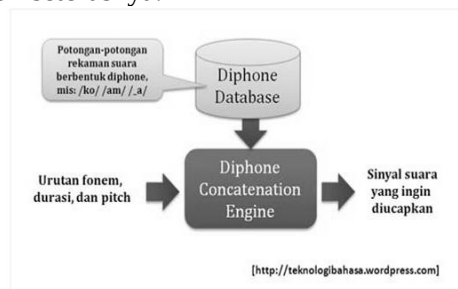
Sistem *Text to Speech* pada prinsipnya terdiri dari dua sub sistem, yaitu: Bagian Konverter Teks ke Fonem (*Text to Phoneme*), serta Bagian Konverter Fonem to Ucapan (*Phoneme to Speech*). Bagian Konverter *text* ke Fonem berfungsi untuk mengubah kalimat masukan dalam suatu bahasa tertentu yang berbentuk teks menjadi rangkaian kode-kode bunyi yang biasanya direpresentasikan dengan kode fonem, durasi serta *pitch*-nya.

Bagian Konverter Fonem ke Ucapan akan menerima masukan berupa kode-kode fonem serta *pitch* dan durasi yang dihasilkan oleh bagian sebelumnya. Berdasarkan kode-kode tersebut, bagian Konverter Fonem ke Ucapan akan menghasilkan bunyi atau sinyal ucapan yang sesuai dengan kalimat yang ingin diucapkan.

2.11 Diphone Concatenation

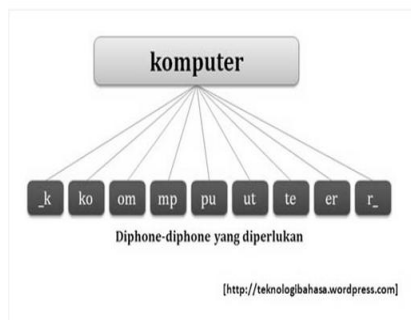
Teknik *diphone concatenation* bekerja dengan cara menggabung-gabungkan segmen-segmen bunyi yang telah direkam sebelumnya. Setiap segmen berupa *diphone* (gabungan dua buah fonem). *Synthesizer* jenis ini dapat menghasilkan bunyi ucapan dengan tingkat kealamian (*naturalness*) yang tinggi.

Pembentukan ucapan pada pensintesa ucapan menggunakan metode *diphone concatenation* pada prinsipnya dilakukan dengan cara menyusun sejumlah *diphone* yang bersesuaian sehingga diperoleh ucapan yang diinginkan. Sebagai contoh, pada gambar diperlihatkan pembentukan kata atau ucapan “komputer” yang disusun dari diphone-diphone /_k/, /ko/, /om/ dan seterusnya.



Gambar 2.1 Blok Diagram Pembentukan Ucapan “komputer” dari Diphone

Supaya pensintesa ucapan dapat mengucapkan semua kemungkinan kata atau kalimat yang ada dalam suatu bahasa, sehingga sistem tersebut harus didukung oleh *diphone database* yang terdiri dari semua kombinasi *diphone* yang ada dalam bahasa tersebut. *Diphone concatenation engine* atau unit pemroses *diphone* akan menerima masukan berupa daftar *fonem* yang ingin diucapkan, masing-masing disertai oleh durasi pengucapannya, serta *pitch* atau frekuensinya. Berdasarkan daftar fonem yang diterima, unit ini akan menentukan susunan *diphone* yang sesuai. Selanjutnya, unit ini akan melakukan smoothing sambungan antar *diphone*, melakukan manipulasi durasi pengucapan serta manipulasi *pitch* (lihat Gambar di bawah). Pada akhirnya, *diphone concatenation engine* akan menghasilkan sinyal ucapan yang sesuai.



Gambar 2.2 Pembentukan Ucapan “komputer” dari Diphone.

2.11 Tuna Netra

Tunanetra adalah istilah umum yang digunakan untuk kondisi seseorang yang mengalami gangguan atau hambatan dalam indra penglihatannya. Berdasarkan tingkat gangguannya/kecacatannya Tunanetra dibagi dua yaitu buta total (*total blind*) dan yang masih mempunyai sisa penglihatan (*Low Visioan*). Alat bantu untuk mobilitasnya bagi tuna netra dengan menggunakan tongkat khusus, yaitu berwarna putih dengan ada garis merah horizontal. Akibat hilang/berkurangnya fungsi indra penglihatannya maka tunanetra berusaha memaksimalkan fungsi indra-indra yang lainnya seperti, perabaan, penciuman, pendengaran, dan lain sebagainya sehingga tidak sedikit penyandang tuna netra yang memiliki kemampuan luar biasa misalnya di bidang musik atau ilmu pengetahuan. Dengan program ini penulis memberi kemudahan kepada para tuna netra agar bisa mendengarkan tanpa bantuan seseorang untuk membacakannya.

III: TUJUAN DAN MANFAAT PENELITIAN

3.1 Tujuan

Tujuan penelitian ini adalah bagaimana membuat *text to speech* menggunakan metode *diphone* concatenation yang dapat digunakan oleh tuna netra.

3.2 Manfaat Penelitian

Adapun penelitian ini diharapkan memberikan manfaat bagi pihak-pihak yang terkait dengan penelitian ini, antara lain:

IV IMPLEMENTASI PROGRAM

4.1 Form Utama

Tampilan utama dari aplikasi Audio Learning menggunakan metode Diphone Concatenation terlihat pada gambar 5.1. berikut ini:



Gambar 4.1. Tampilan Utama Aplikasi

4.2 Form Pilih File

Pada menu PDF merupakan pilihan yang mengarahkan pada tampilan yang berekstensi .PDF, pengguna bisa membuka langsung file yang berekstensi *.PDF pada pilihan menu PDF. Kemudian bisa mengcopykan paragraph yang akan diproses menjadi *speech* ke menu SPEAKING. Pada menu Speaking merupakan menu yang menampilkan teks-teks yang diproses menjadi *speech*. Ketika teks-teks tersebut telah diproses menjadi *speech* pengguna bisa langsung mendengarkan isi yang terdapat dalam teks tersebut.

Text to speech dirancang menggunakan bahasa Indonesia memudahkan pengguna yang mengerti berbahasa Indonesia. Karena *prosody* di setiap bahasa berbeda-beda, ketika program *text to speech* menggunakan bahasa Inggris atau *prosody* Inggris tetapi isi dari teks menggunakan bahasa Indonesia maka *speech* yang akan didengar adalah bahasa Indonesia yang berprosody Inggris. Hasil tersebut akan membingungkan pengguna untuk memahami isi teks tersebut.

Ketika proses speech berlangsung terdapat status *speaking* yang menunjukkan bahwa teks sedang dibacakan. Apabila proses speech telah selesai maka status yang ditunjukkan yaitu diam, dan status *not speaking* menunjukkan bahwa dalam memo tidak terdapat teks yang akan di proses menjadi speech. Speech yang dibacakan bisa diatur kecepatannya dalam membacakan teks dengan mengatur *trackbar* yang sudah tersedia dalam aplikasi ini. Pada tombol katakana dan stop merupakan tombol untuk proses *speech* dan menghentikan *speech*.



Gambar 4.3. Tampilan Text to Speech

DAFTAR PUSTAKA

- [1] Dutoit, Thierry (1996), High-Quality Text-to-Speech Synthesis : an Overview, *Journal of Electrical & Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis*, vol. 17 n°1, pp. 25-37.
- [2] Haykin, Simon (1994). *Neural Network : A Comprehensive Foundation*, Macmillan Publishing Company.
- [3] Moulines E., F. Charpentier (1990), Pitch Synchronous Waveform Processing technique for text-to-speech Synthesis Using Diphones, *Speech Communication*, August, 453-467.
- [4] JR, John R. Deller, John G. Proakis, John H.L. Hansen (1993), *Discrete-Time Processing of Speech Signals*, Macmillan Inc.
- [5] L. Rabiner, R. Schafer (1978), *Digital Signal Processing of Speech Signal*, Signal Processing, Prentice Hall.
- [6] Shaw-Hwa Hwang and Sin-Horng Chen (1995), "A Prosodic Model of Mandarin Speech And Its application to Pitch Level Generation for Text-to-Speech", *IEEE*.
- [7] Sin-Horng Chen, Shaw-Hwa Hwang and Yih-Ru Wang (May 1998),"An RNN-Based Prosodic Information Synthesizer for Mandarin Text-to-Speech, *IEEE*, Vol 6 No.3
- [8] Vainio, Marti (2001), *Artificial Neural Network Based Prosody Models for Finnish Text-to-Speech Synthesis*, University of Helsinki.
- [9] Y. Sagisaka,"On the prediction of Global F0 shape for Japanese text-to-speech,"*ICASSP*, pp.325-328, 1990