



SPEECH RECOGNITION

Pertemuan Ke-3
Semester Genap 2016-2017

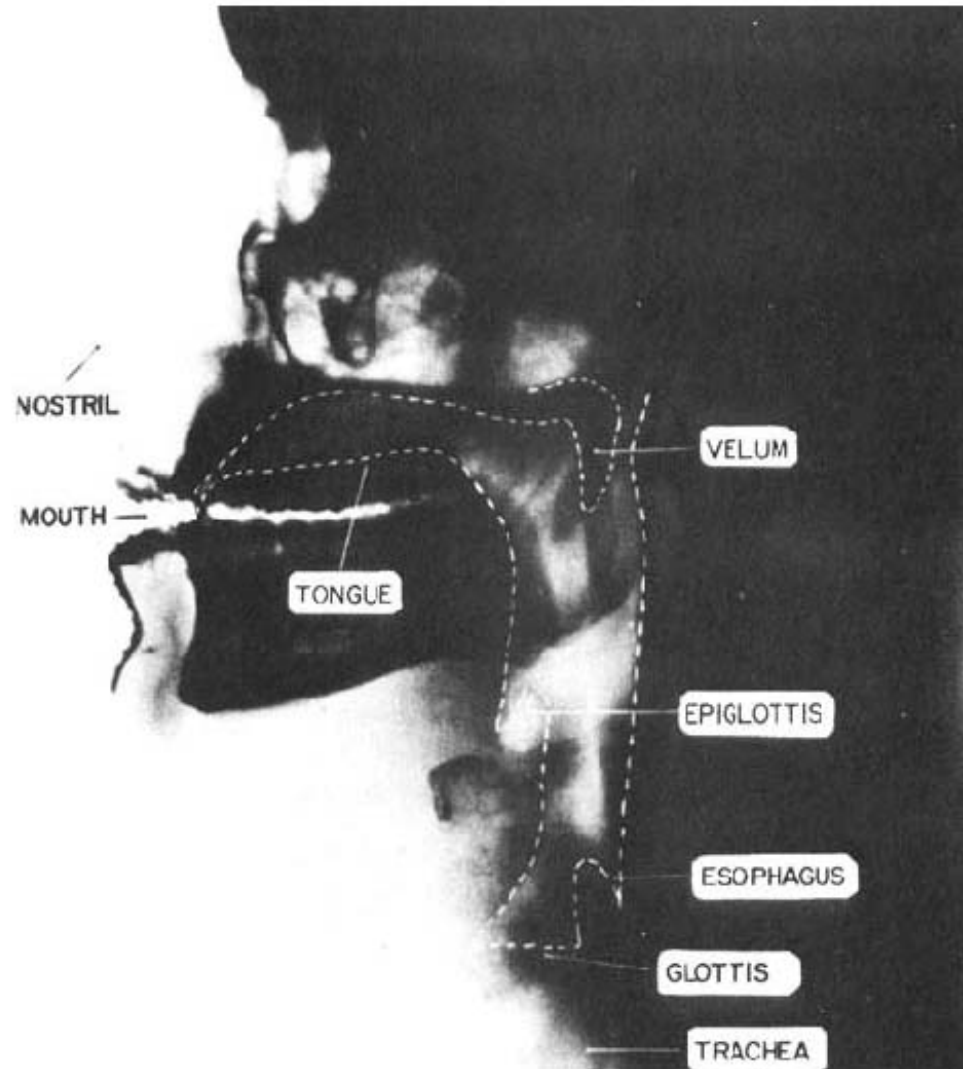
PROSES PEMBENTUKAN DAN KARAKTERISTIK SINYAL UCAPAN

Dr. R. Rizal Isnanto, ST, MM, MT
Andi Widiasmoro, ST, M.Kom

1. Sistem Pembentukan Ucapan

- Ucapan manusia dihasilkan oleh suatu sistem produksi ucapan yang dibentuk oleh alat-alat ucap manusia.
- Proses tersebut dimulai dengan formulasi pesan dalam otak pembicara.
- Pesan tersebut akan diubah menjadi perintah-perintah yang diberikan kepada alat-alat ucap manusia, sehingga akhirnya dihasilkan ucapan yang sesuai dengan pesan yang ingin diucapkan.

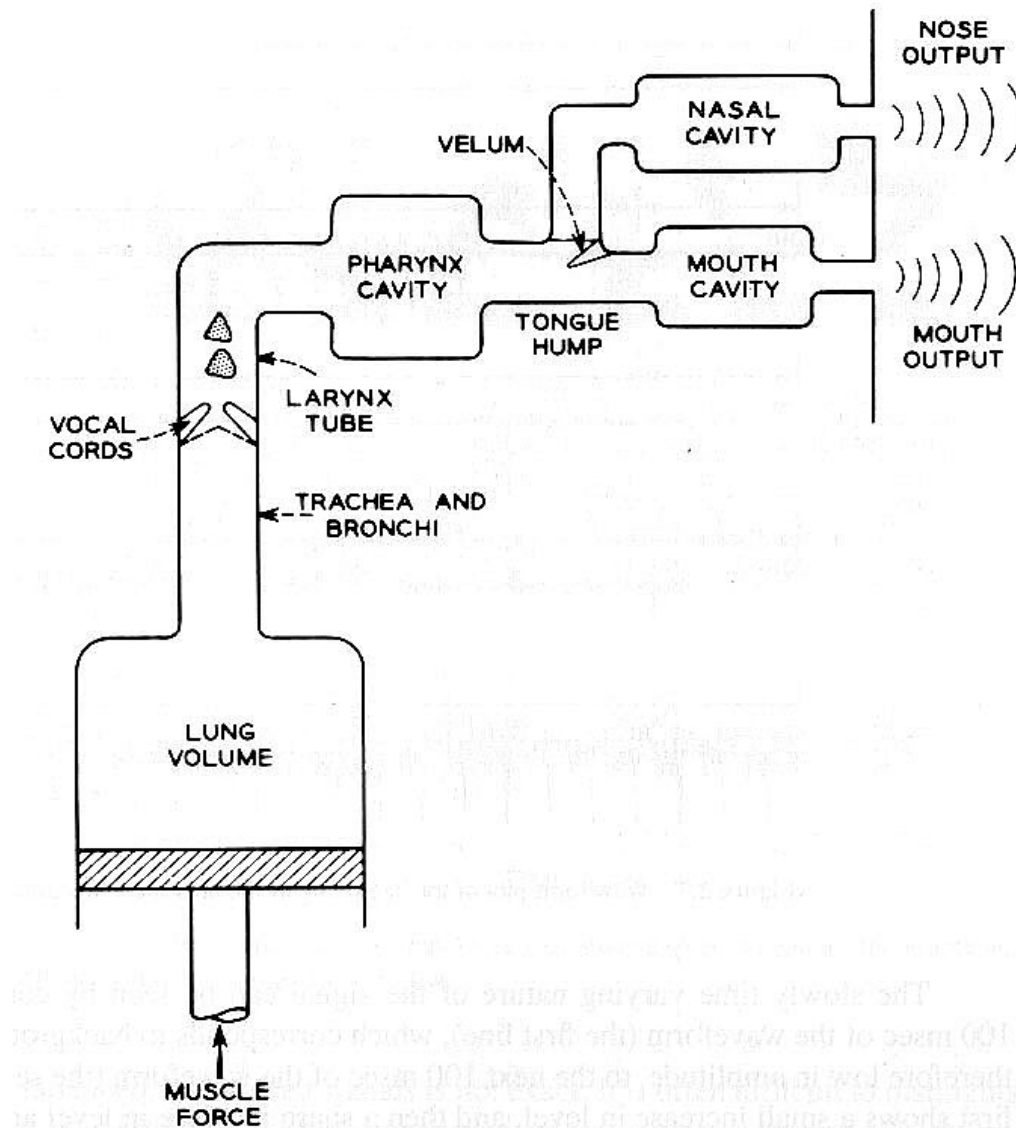
Foto Sinar X Penampang Alat-Alat Ucap Manusia



Keterangan tentang Alat-alat Ucap

- Gambar memperlihatkan foto sinar X penampang alat-alat ucap manusia. *Vocal tract* pada gambar tersebut ditandai oleh garis putus-putus, dimulai dari *vocal cords* (pita suara) atau glottis, dan berakhir pada mulut.
- *Vocal tract* (saluran suara) terdiri dari *pharynx* (tekak/tenggorokan) yang merupakan koneksi antara *esophagus* (kerongkongan) dengan mulut, dan mulut. Panjang *vocal tract* pria pada umumnya sekitar 17 cm dan wanita 14 cm
- Daerah pertemuan *vocal tract* ditentukan oleh lidah, bibir, rahang, dan bagian belakang langit-langit; luasnya berkisar antara 20 cm² sampai dengan mendekati nol (ketika menutup)
- *Nasal tract* (saluran hidung) mulai dari bagian belakang langit-langit dan berakhir pada *nostrils* (lubang hidung).
- Pada keadaan tertentu, suara nasal (suara sengau) akan dikeluarkan melalui rongga ini.

Model Sistem Produksi Ucapan Manusia



Keterangan tentang model sistem produksi ucapan manusia

- Gambar memperlihatkan model sistem produksi ucapan manusia yang disederhanakan.
- Pembentukan ucapan dimulai dengan adanya hembusan udara yang dihasilkan oleh paru-paru.
- Cara kerjanya mirip seperti piston atau pompa yang ditekan untuk menghasilkan tekanan udara. Pada saat *vocal cord* berada dalam keadaan tegang, aliran udara akan menyebabkan terjadinya vibrasi pada *vocal cord* dan menghasilkan bunyi ucapan yang disebut *voiced speech sound*.
- Pada saat *vocal cord* berada dalam keadaan lemas, aliran udara akan melalui daerah yang sempit pada *vocal tract* dan menyebabkan terjadinya *turbulensi*, sehingga menghasilkan suara yang dikenal sebagai *unvoiced sound*.

Mengapa ucapan menjadi unik?

- Ucapan dihasilkan sebagai rangkaian atau urutan komponen-komponen bunyi-bunyi pembentuknya.
- Setiap komponen bunyi yang berbeda dibentuk oleh perbedaan **posisi**, **bentuk**, serta **ukuran** dari alat-alat ucap manusia yang berubah-ubah selama terjadinya proses produksi ucapan.


2. Representasi Sinyal Ucapan

- Sinyal ucapan merupakan sinyal yang berubah terhadap waktu dengan kecepatan perubahan yang relatif lambat.
- Jika diamati pada selang waktu yang pendek (antara 5 sampai dengan 100 milidetik), karakteristiknya praktis bersifat tetap
- Namun, jika diamati pada selang waktu yang lebih panjang karakteristiknya terlihat berubah-ubah sesuai dengan kalimat yang sedang diucapkan.

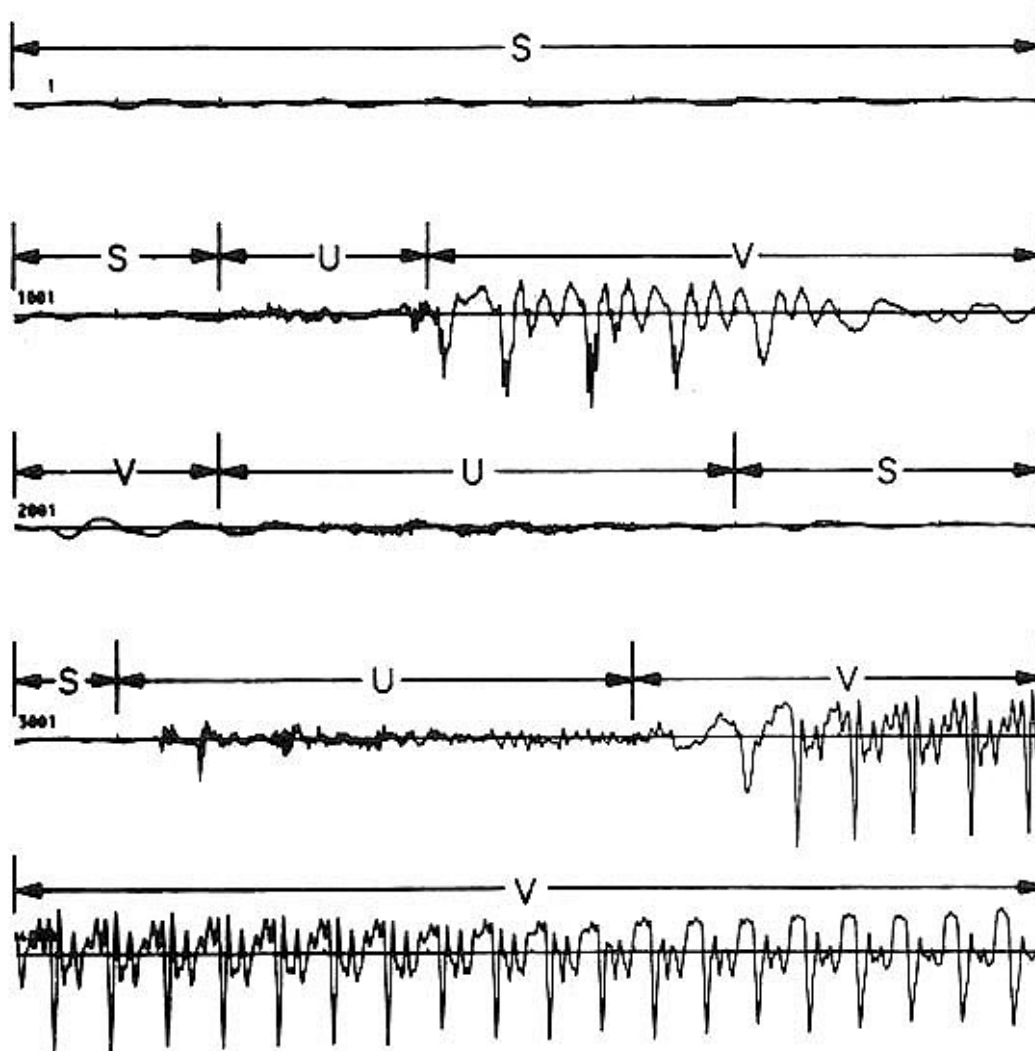
Komponen Sinyal Ucapan

- Ada berbagai cara untuk mengklasifikasikan bagian-bagian atau komponen sinyal ucapan.
- Salah satu cara yang sederhana adalah dengan cara mengklasifikasikannya menjadi tiga keadaan yang berbeda, yaitu
 - (1) **silence (S)**, keadaan pada saat tidak ada ucapan yang diucapkan;
 - (2) **unvoiced (U)**, keadaan pada saat pita suara tidak melakukan vibrasi, sehingga suara yang dihasilkan bersifat tidak periodik atau bersifat acak
 - (3) **voiced (V)**, keadaan pada saat terjadinya vibrasi pada pita suara, sehingga menghasilkan suara

Contoh sinyal ucapan “It’s time”

- Contoh sinyal ucapan dari suatu kalimat bahasa Inggris “*It’s time*” yang diucapkan oleh seorang pria pada slide berikut.
- Setiap baris pada gambar tersebut memperlihatkan potongan sinyal selama 100ms, sehingga seluruh gambar tersebut memperlihatkan sinyal ucapan sepanjang 500ms
- It’s Time 

Contoh sinyal ucapan "It's time"



Penjelasan (1)

- Pada gambar di atas sudah tercantum label-label S, U dan V yang dapat mempermudah untuk mengamati perbedaan keadaan-keadaan tersebut.
- Baris pertama serta awal baris kedua ditandai dengan S, artinya bagian tersebut merepresentasikan keadaan diam dimana pembicara belum mengucapkan apapun. *Amplitude* kecil yang tampak pada periode tersebut merupakan derau latar belakang yang ikut terekam
- Suatu periode singkat *unvoiced* (U) tampak mendahului vokal pertama dalam kata "It".
- Selanjutnya diikuti oleh daerah *voiced* (V) yang cukup panjang, merepresentasikan vokal "i".
- Berikutnya diikuti oleh daerah *unvoiced* (U) yang merepresentasikan daerah pelemahan pengucapan "i".
- Setelah itu diikuti oleh *silence* (S) yang merupakan bagian dari fonem "t", dan seterusnya.

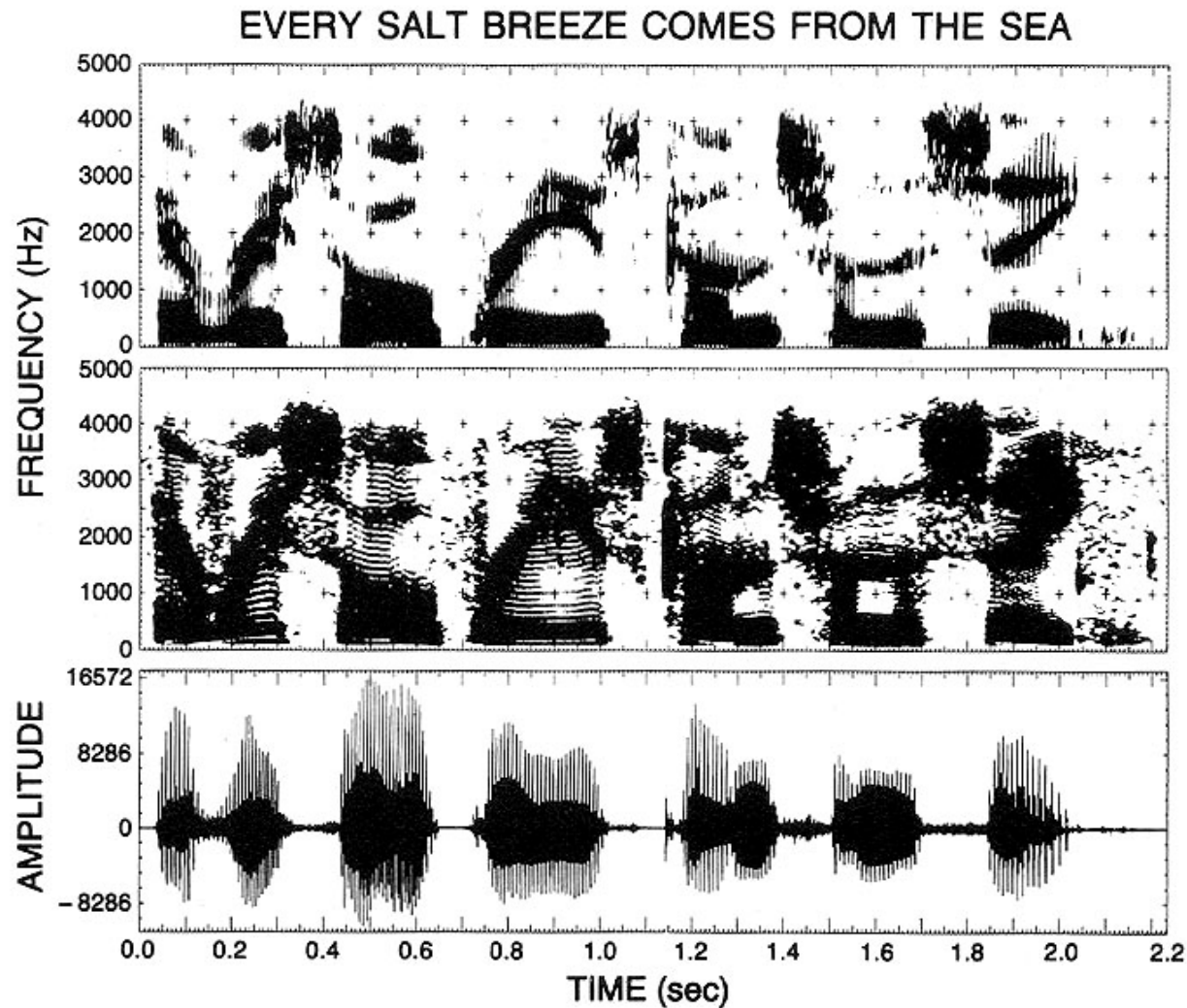
Penjelasan (2)

- Dari contoh tersebut jelas bahwa segmentasi ucapan menjadi S, U dan V tidak bersifat eksak, artinya ada daerah-daerah yang tidak dapat dikategorikan dengan tegas ke dalam salah satu dari tiga kategori tersebut.
- Salah satu penyebabnya adalah perubahan dari keadaan-keadaan alat ucap manusia yang tidak bersifat diskret dari satu keadaan ke keadaan lainnya, sehingga bunyi transisi dari satu segmen ke segmen lainnya menghasilkan bentuk yang tidak mudah ditentukan.
- Selain itu, ada segmen-segmen ucapan yang mirip atau bahkan mengandung *silence* di dalamnya.
- Representasi sinyal dalam diagram waktu terhadap amplitude seperti gambar tersebut seringkali tidak cukup untuk mendapatkan besaran-besaran kuantitatif yang efektif untuk melakukan analisis dari suatu ucapan.

SPEKTOGRAM

- Untuk melakukan analisis sinyal ucapan, lebih sering digunakan representasi spektral menggunakan spektogram seperti terlihat pada slide berikut.
- Dengan menggunakan spektogram, dapat diidentifikasi komponen-komponen frekuensi dari suatu segmen ucapan.
- Segmen ucapan yang bentuknya mirip pada domain waktu, lebih mudah dibedakan pada spektogram dengan cara melihat perbedaan komponen frekuensinya.

Spektogram Pita Lebar, Spektogram Pita Sempit, dan Amplitudo Ucapan dari suatu Kalimat



Penjelasan

- Spektogram dibedakan menjadi spektogram pita lebar (*wideband spectrogram*) dan spektogram pita sempit (*narrowband spectrogram*).
- Spektogram pita lebar adalah analisis spektral pada suatu interval sepanjang 15ms menggunakan filter dengan lebar pita 125 Hz serta analisis detail yang dilakukan setiap 1 ms.
- Spektogram pita sempit adalah analisis spectral pada suatu interval sepanjang 50 ms menggunakan filter dengan lebar pita 40 Hz serta analisis detail yang dilakukan setiap 1 ms.
- Spektogram pita lebar dapat digunakan untuk melihat komponen-komponen frekuensi utama dari suatu ucapan dengan jelas, seperti terlihat pada gambar paling atas dari gambar tersebut.
- Sebagian komponen frekuensi yang tidak dominan menjadi tidak terlihat pada spektogram pita lebar.
- Untuk melihat komponen-komponen frekuensi yang lebih rinci dilakukan menggunakan spektogram pita sempit, seperti yang terlihat pada gambar kedua dari atas pada gambar tersebut

Implementasi Analisis Spektral

- Dalam kegiatan penelitian dan pengembangan sistem **TTS (*text-to-speech*)**, analisis spektral di antaranya digunakan untuk melakukan segmentasi komponen-komponen sinyal ucapan, identifikasi komponen frekuensi segmen ucapan, serta analisis frekuensi dasar yang diperlukan untuk **analisis intonasi ucapan**.

3. Karakteristik Sinyal Ucapan

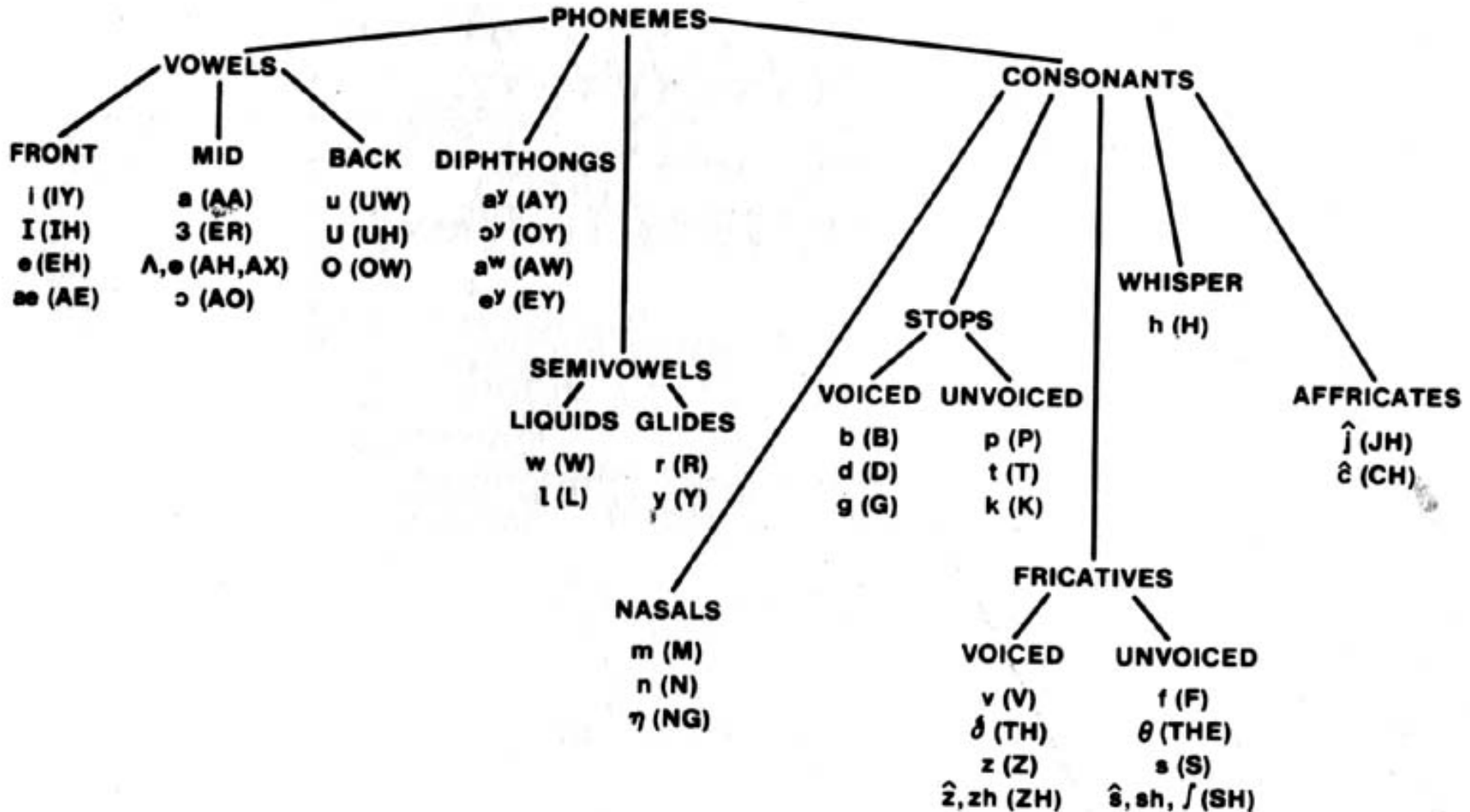
- Unit bunyi terkecil yang dapat dibedakan oleh manusia disebut fonem. Suatu ucapan kata atau kalimat pada prinsipnya dapat dilihat sebagai urutan fonem. Himpunan fonem yang ada dalam suatu bahasa berbeda-beda. Setiap fonem disimbolkan dengan suatu simbol yang unik.
- Saat ini ada beberapa standar cara penamaan fonem yang berlaku di antaranya: adalah standar:
 - (1) IPA (International Phonetic Alphabet)1,
 - (2) ARPABET, serta
 - (3) SAMPA
- Tabel berikut memperlihatkan daftar fonem bahasa Inggris-Amerika serta representasinya dalam simbol-simbol IPA serta ARPABET

Fonem-fonem Bahasa Inggris-Amerika dalam standar IPA dan ARPABET

TABLE 2.1. A condensed list of phonetic symbols for American English.

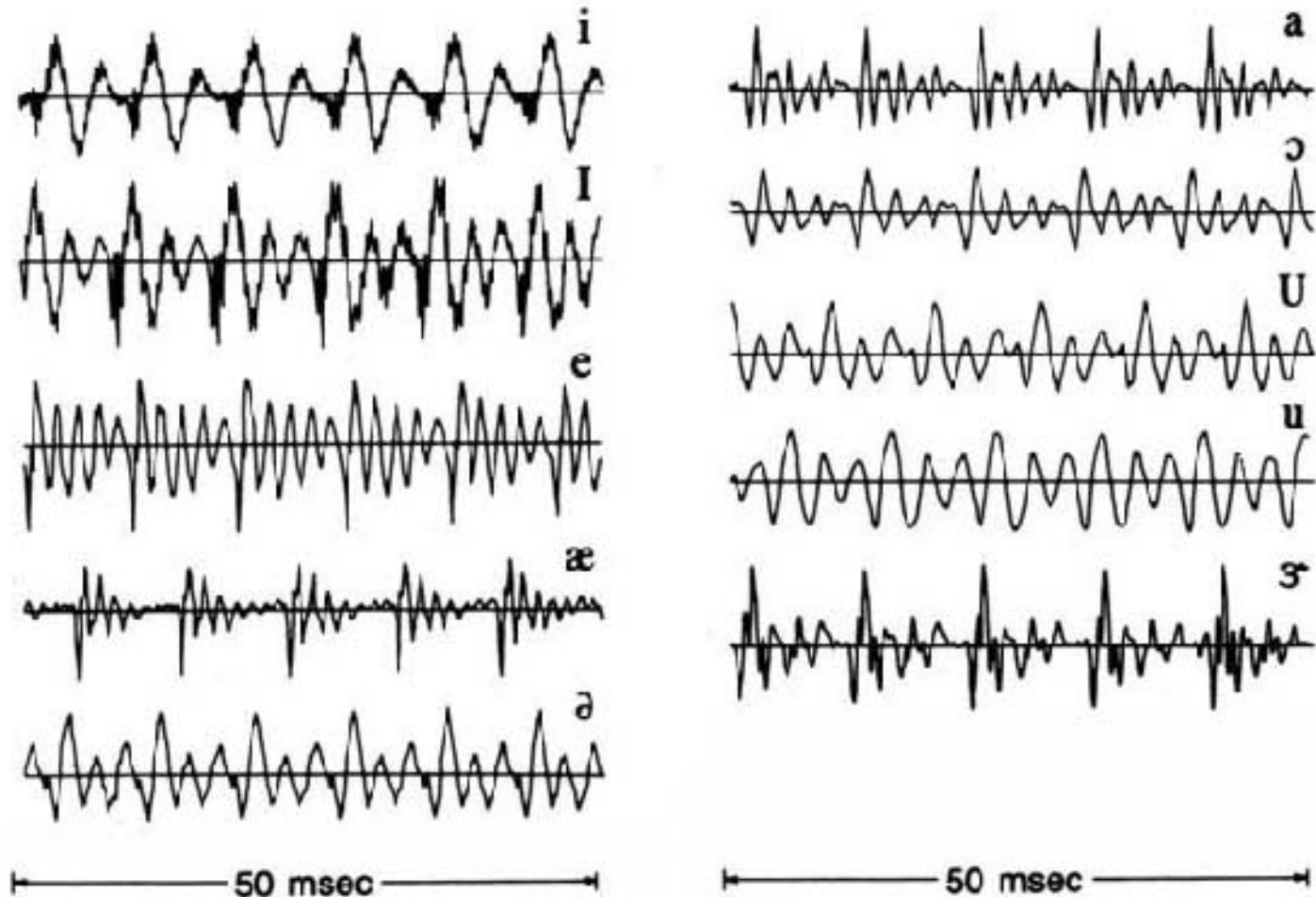
Phoneme	ARPABET	Example	Phoneme	ARPABET	Example
/i/	IY	<u>beat</u>	/ŋ/	NX	<u>sing</u>
/ɪ/	IH	<u>bit</u>	/p/	P	<u>pet</u>
/e/ (eʏ)	EY	<u>bait</u>	/t/	T	<u>ten</u>
/ɛ/	EH	<u>bet</u>	/k/	K	<u>kit</u>
/æ/	AE	<u>bat</u>	/b/	B	<u>bet</u>
/ɑ/	AA	<u>Bob</u>	/d/	D	<u>debt</u>
/ʌ/	AH	<u>but</u>	/g/	H	<u>get</u>
/ɔ/	AO	<u>bought</u>	/h/	HH	<u>hat</u>
/o/ (oʷ)	OW	<u>boat</u>	/f/	F	<u>fat</u>
/ʊ/	UH	<u>book</u>	/θ/	TH	<u>thing</u>
/u/	UW	<u>boot</u>	/s/	S	<u>sat</u>
/ə/	AX	<u>about</u>	/ʃ/ (sh)	SH	<u>shut</u>
/ɪ/	IX	<u>roses</u>	/v/	V	<u>vat</u>
/ɜ/	ER	<u>bird</u>	/ð/	DH	<u>that</u>
/ɔ/	AXR	<u>butter</u>	/z/	Z	<u>zoo</u>
/aʷ/	AW	<u>down</u>	/ʒ/ (zh)	ZH	<u>azure</u>
/aʏ/	AY	<u>buy</u>	/tʃ/ (tsh)	CH	<u>church</u>
/ɔʏ/	OY	<u>boy</u>	/dʒ/ (dzh, j)	JH	<u>judge</u>
/y/	Y	<u>you</u>	/w/	WH	<u>which</u>
/w/	W	<u>wit</u>	/l/	EL	<u>battle</u>
/r/	R	<u>rent</u>	/m/	EM	<u>bottom</u>
/l/	L	<u>let</u>	/n/	EN	<u>button</u>
/m/	M	<u>met</u>	/ŋ/	DX	<u>batter</u>
/n/	N	<u>net</u>	/ʔ/	Q	(glottal stop)

Daftar fonem serta pengkategorisasiannya untuk bahasa Inggris-Amerika

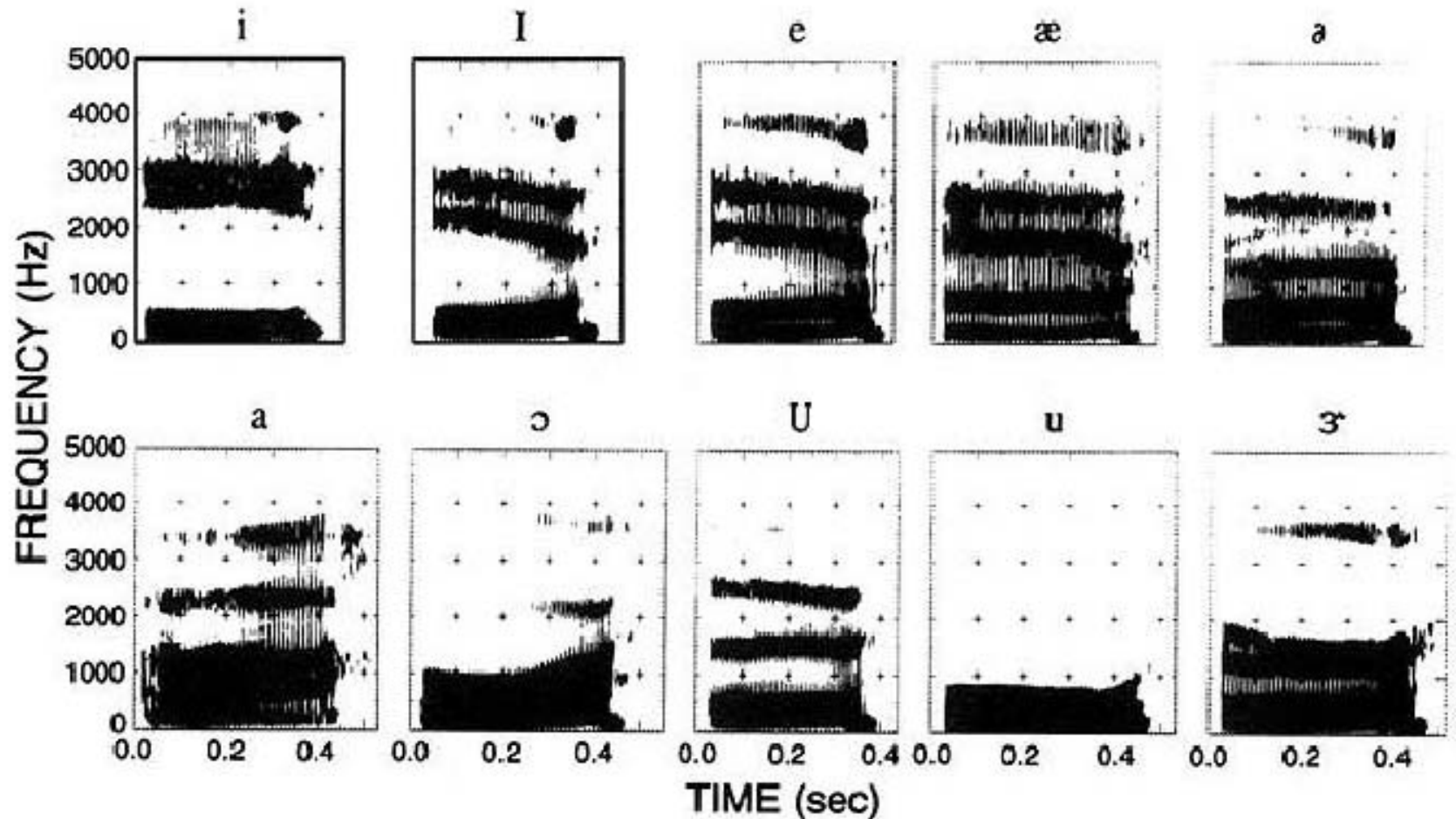


4. Vokal

- Sinyal ucapan vokal memiliki bentuk kuasi periodik
- Contoh : Bentuk Sinyal Ucapan Vokal Bahasa Inggris



- Setiap vokal mempunyai komponen frekuensi tertentu yang membedakan karakter satu fonem vokal dengan fonem vokal lainnya
- Contoh : Spektogram Sinyal Ucapan Vokal Bahasa Inggris



- Fonem vokal Bahasa Inggris mencakup fonem-fonem /IY/, /IH/, /EH/, /AE/, /AA/, /ER/, /AH/, /AX/, /AO/, /UW/, /UH/, dan /OW/
- Dalam Bahasa Indonesia
 - Vokal Dasar : a, i, u, e, o
 - Vokal Gabungan / diftong : au, ai, oi

Consonants



Vowels

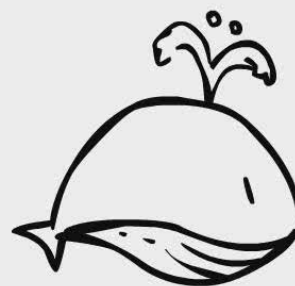


Diphthongs



whale

/weɪl/



Learn ► Enjoy ► Succeed

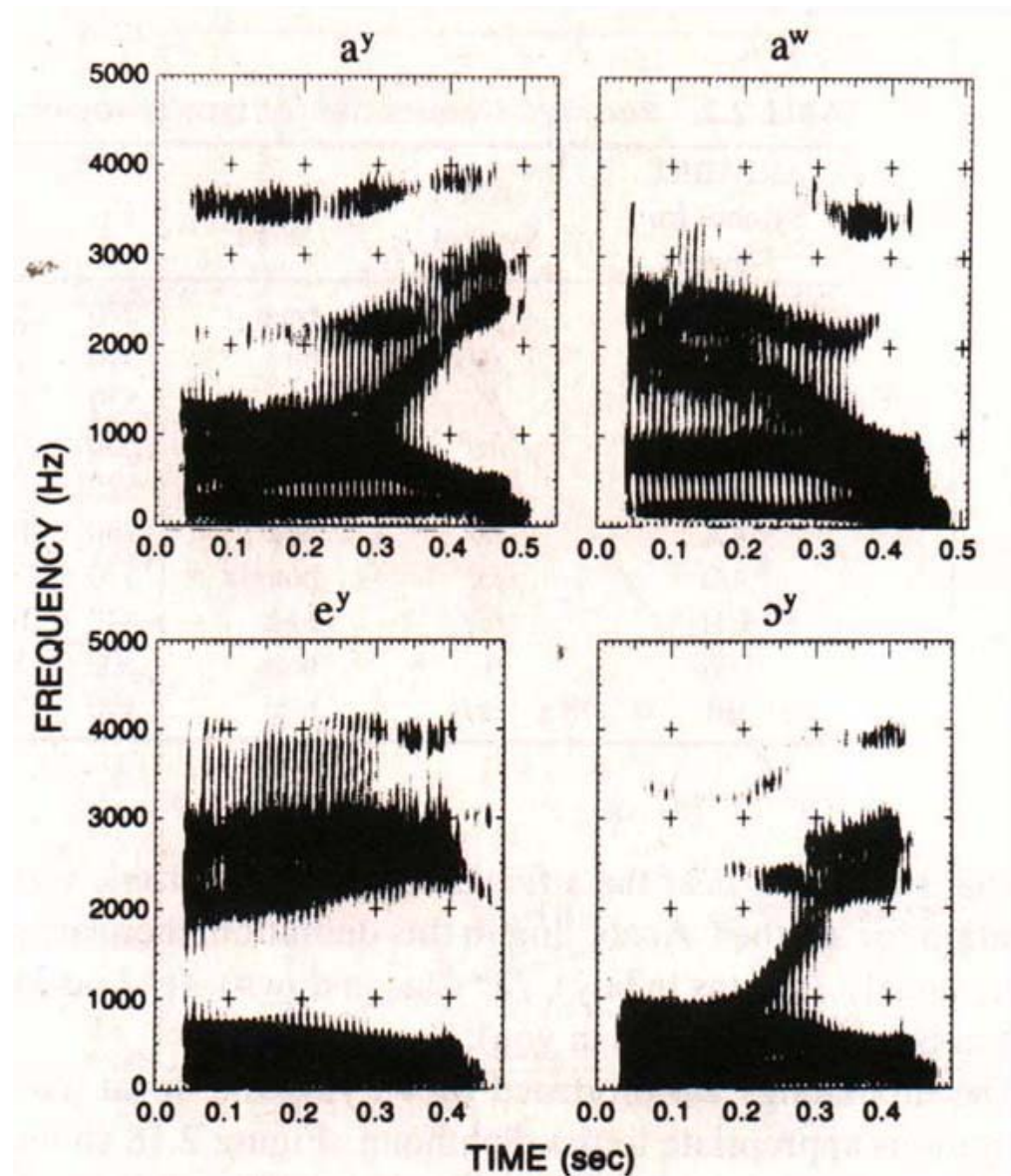
www.cambridgeenglishonline.com

Cambridge English Online

5. Diftong

- Diftong pada prinsipnya adalah dua fonem vokal yang berurutan dan diucapkan tanpa jeda.
- Fonem diftong Bahasa Inggris mencakup /AY/, /OY/, /AW/, dan /EY/. Karakteristik diftong mirip dengan karakteristik fonem-fonem vokal pembentuknya disertai bentuk transisinya.
- Dalam Bahasa Indonesia, dikenal diftong-diftong: /ai/ → pantai; /oi/ → amboi; /au/ → lampau; /ei/ → eit!

Spektrogram Sinyal Ucapan Diftong Bahasa Inggris



Consonants

p	b	t	d
k	g	m	n
ŋ	f	v	θ
ð	s	z	ʃ
ʒ	h	tʃ	dʒ
r	l	j	w

Vowels

iː	ɪ	e	æ
ɑː	ʌ	ɒ	ɔː
ʊ	uː	ʊː	ə

Diphthongs

eɪ	aɪ	ɔɪ	əʊ
aʊ	eə	ɪə	ʊə

father

/fɑːðə/



Learn ► Enjoy ► Succeed

www.cambridgeenglishonline.com

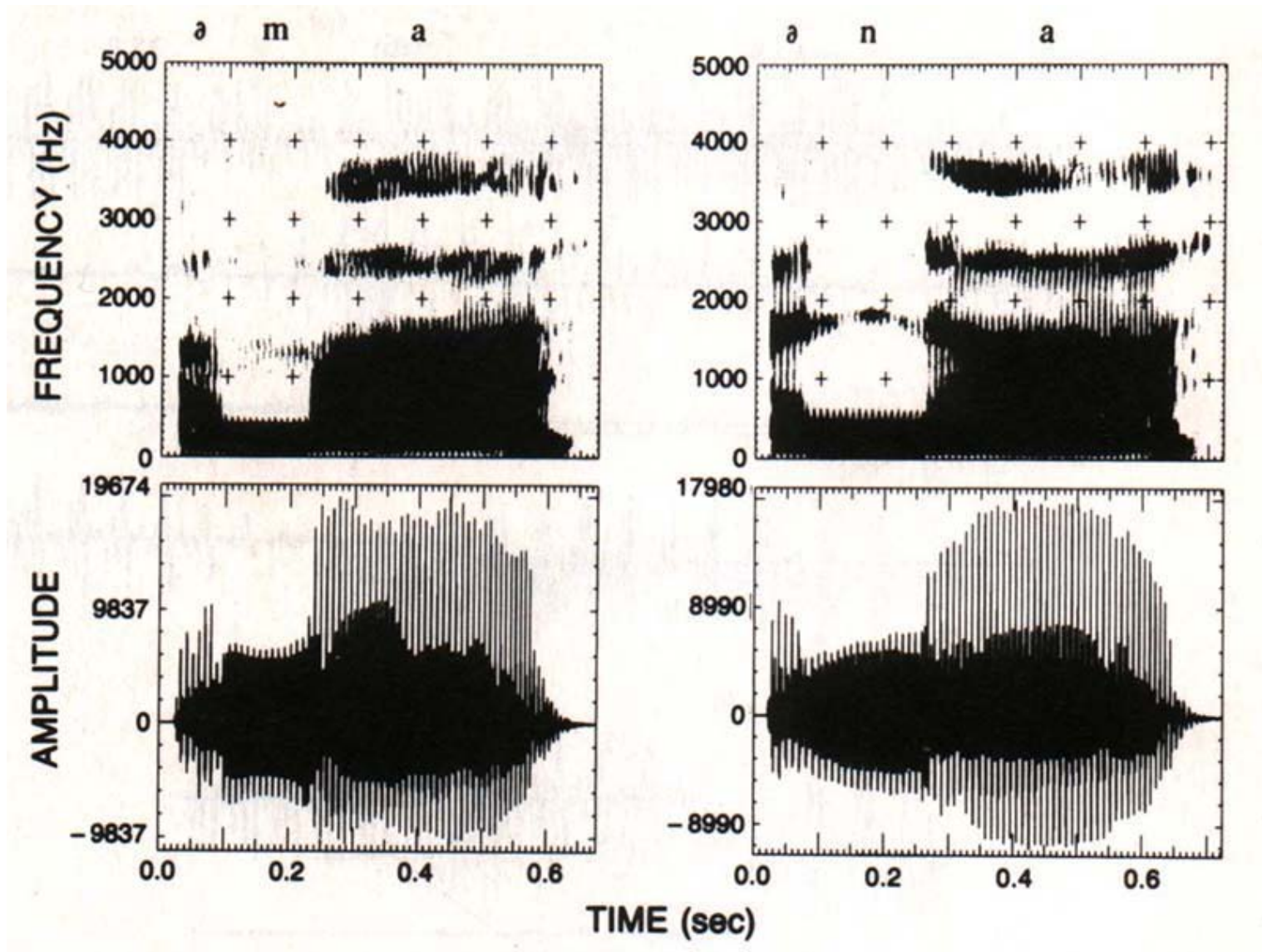
Cambridge English Online



6. Konsonan Nasal

- Konsonan nasal dibangkitkan dengan eksitasi glotal dan *vocal tract* mengerut total pada beberapa titik tertentu sepanjang lintasan pengucapan
- Bagian belakang langit-langit merendah, sehingga udara mengalir melalui *nasal tract* dengan suara yang dipancarkan melalui lubang hidung
- Konsonan nasal Bahasa Inggris adalah /M/, /N/, dan /NX/ = 'ng'.

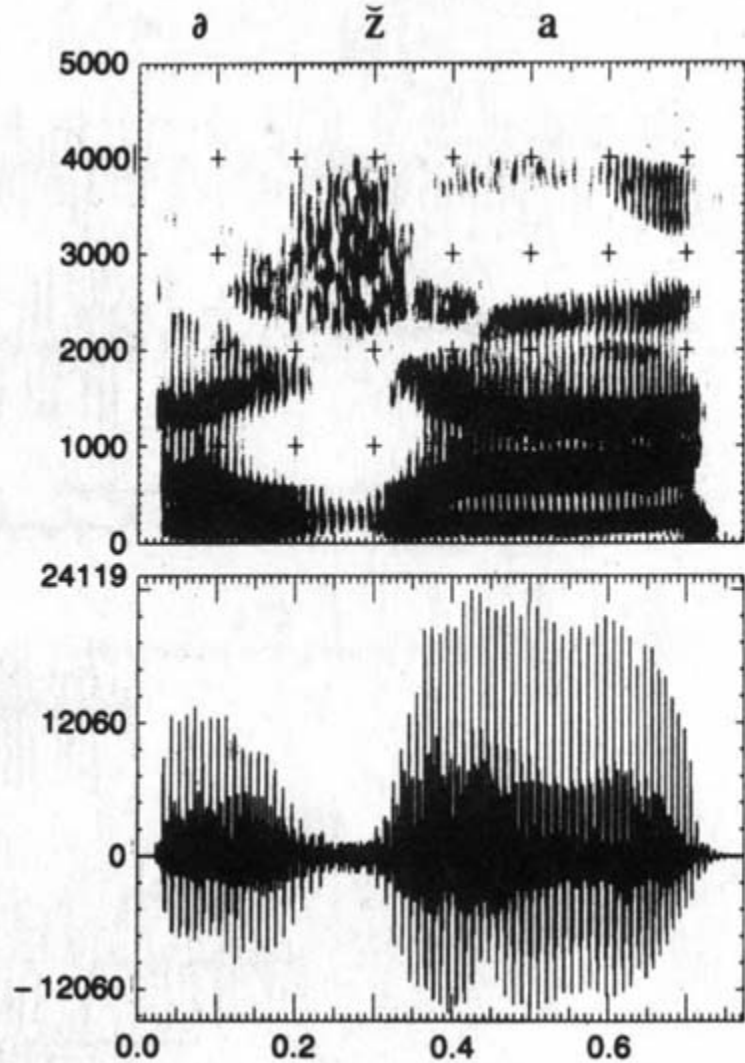
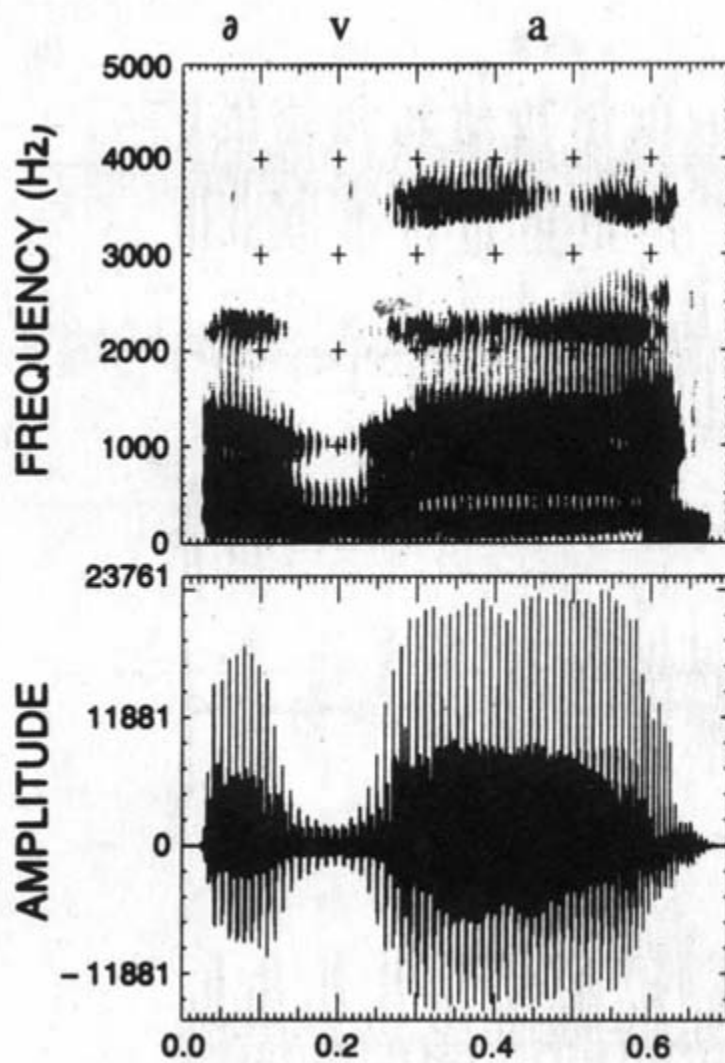
Sinyal dan Spektrogram Konsonan Nasal Bahasa Inggris



7. Konsonan Frikatif

- Konsonen frikatif pada prinsipnya dapat dibedakan menjadi frikatif *unvoiced* serta *voiced*. Fonem Bahasa Inggris yang termasuk frikatif *unvoiced* adalah /F/, /TH/, /S/, dan /SH/, sedangkan yang termasuk frikatif *voiced* adalah /V/, /Z/, = 'zoo' dan /ZH/ = 'azure'
- Frikatif *unvoiced* dibentuk dengan suatu eksitasi terhadap *vocal tract* dengan suatu aliran udara yang tetap, sehingga menyebabkan turbulensi di daerah yang mengkerut dalam *vocal tract*.
- Frikatif *voiced* agak berbeda dengan frikatif *unvoiced*. Pada frikatif *voiced*, suara dihasilkan oleh dua sumber eksitasi. Sumber eksitasi lainnya adalah glotis

Sinyal dan Spektrogram Konsonan Frikatif Bahasa Inggris



TIME (sec)

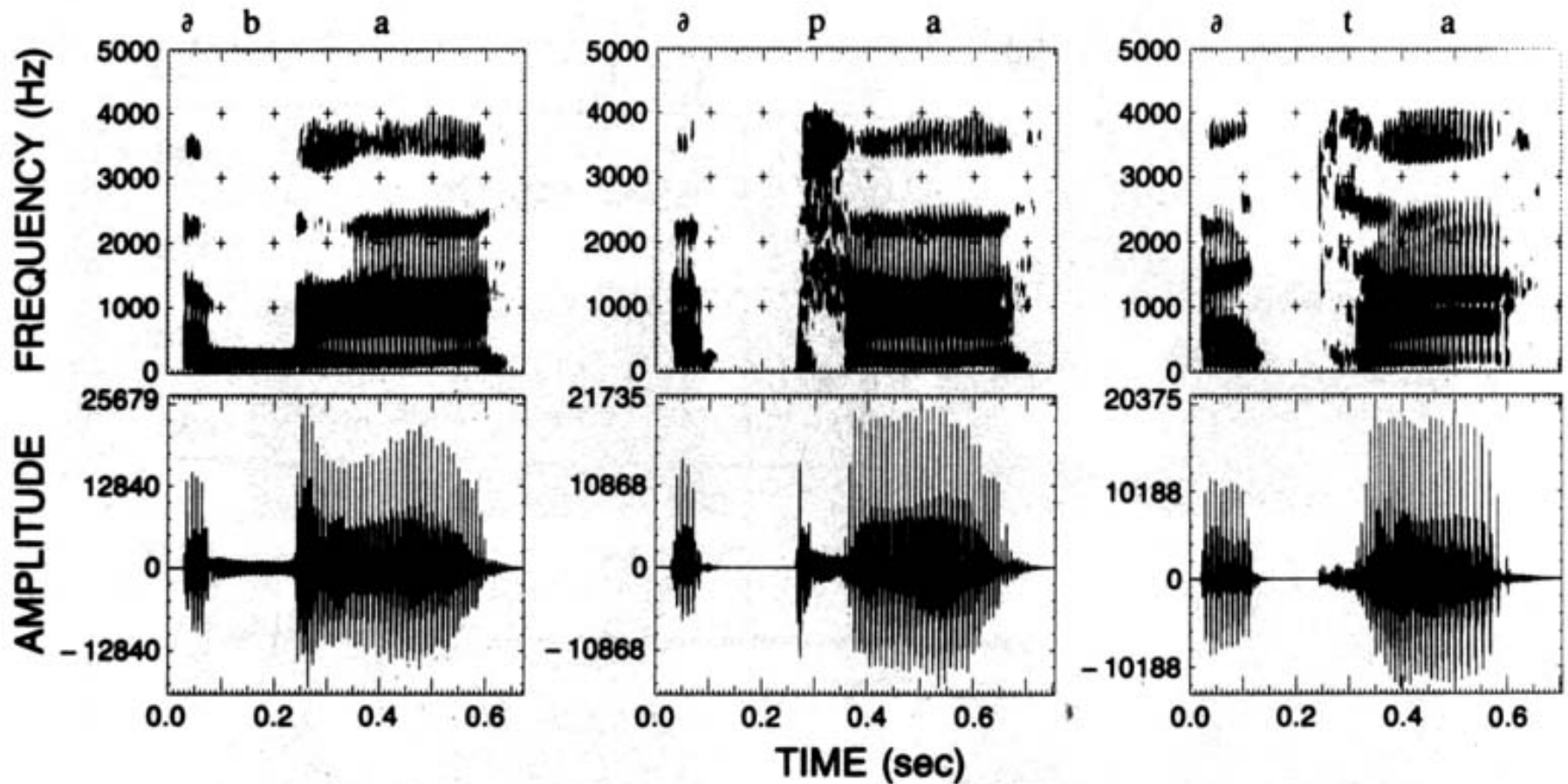
8. Konsonan Stop

- Seperti konsonan frikatif, konsonen stop dapat dibedakan menjadi konsonan stop *unvoiced* serta *voiced*. Pada *Unvoiced*, pita suara dalam keadaan lemas, aliran udara akan melalui daerah yang sempit pada *vocal tract* dan menyebabkan terjadinya turbulensi.
- Konsonan stop memiliki bentuk yang berbeda dengan konsonan-konsonan lainnya. Konsonan ini memperlihatkan pola transient dan tidak kontinu.
- Konsonan ini dibentuk dengan cara memberikan tekanan pada kondisi pengerutan total di bagian rongga mulut tertentu, dan segera diikuti dengan pelepasan.

8. Konsonan Stop...

- Untuk fonem /B/ pengerutan terjadi di bibir,
- Untuk fonem /D/ pengerutan terjadi di belakang gigi depan,
- Sedangkan untuk fonem /G/ pengerutan terjadi di sekitar bagian belakang langit-langit.
- Selama periode total pengerutan terjadi, tidak ada suara yang dikeluarkan dari mulut, sehingga fonem ini selalu mengandung bagian yang menyerupai *silence*.
- Fonem Bahasa Inggris yang termasuk konsonan stop *unvoiced* adalah /P/, /T/, dan /K/, sedangkan yang termasuk konsonan stop *voiced* adalah /B/, /D/, dan /G/.

Sinyal dan Spektrogram Konsonan Stop Bahasa Inggris



Consonants

p	b	t	d
k	g	m	n
ŋ	f	v	θ
ð	s	z	ʃ
ʒ	h	tʃ	dʒ
r	l	j	w

Vowels

i:	ɪ	e	æ
ɑ:	ʌ	ɒ	ɔ:
ʊ	u:	ʊ:	ə

Diphthongs

eɪ	aɪ	ɔɪ	əʊ
aʊ	eə	ɪə	ʊə





TERIMA KASIH