

## ANALISIS VOCAL TRACT PADA KAJIAN AKUSTIK VOKAL BAHASA INDONESIA

Ichwan Suyudi<sup>1</sup>  
Debyo Saptono<sup>2</sup>

Fakultas Sastra Inggris, Universitas Gunadarma  
<sup>1</sup>ichwan,<sup>2</sup>debyo@staff.gunadarma.ac.id

### ABSTRAK

Penelitian ini bertujuan menganalisis karakteristik bunyi ujaran atau disebut analisis formana atau vocal tract. Kontur nada vokal dianalisis hanya apabila frekuensi fundamentalnya menunjukkan pola naik-turun yang jelas. Untuk kepentingan tersebut maka frekuensi fundamental ( $f_0$ ,  $f_1$ , dan  $f_2$ ) dihitung secara tersendiri dalam titik-titik tertentu. Sebagai bahan perbandingan dihitung pula nilai  $f_0$ ,  $f_1$  dan  $f_2$  dalam konteks. Metode yang digunakan adalah perangkat lunak Praat dan aplikasi VisArtico. Hasil penelitian menunjukkan bahwa bunyi vokal /a/, menunjukkan nilai  $f_1$  dan  $f_2$  yang berbeda ketika data ujaran dilakukan melalui dua proses perekaman yang berbeda. Perbedaan ini diduga dihasilkan oleh alofonik yang mempengaruhi produksi bunyi vokal tersebut.

**Kata kunci:** vocal tract, akustik, bunyi vokal, bahasa Indonesia

### PENDAHULUAN

Azhar (2001) dalam penelitiannya, menjelaskan bahwa suara dihasilkan melalui dua buah proses, yaitu generation dan filtering. Proses generation adalah pertama kali bunyi ujaran akan diproduksi melalui bergetarnya pita suara (*vocal cord* dan *vocal fold*) yang berada di larynx untuk menghasilkan bunyi periodik. Bunyi periodik bersifat konstan dan selanjutnya difilter melalui vocal tract atau artikulator yang terdiri atas lidah (*tongue*), gigi (*teeth*), bibir (*lips*), langit-langit (*palate*) dan lain-lain sehingga bunyi tersebut menjadi bunyi keluaran (*output*) berupa bunyi vokal (*vowel*) dan atau bunyi konsonan (*consonant*) yang membentuk kata-kata yang memiliki arti yang nantinya dapat dianalisa untuk pengenalan bunyi ucapan atau sering dikenal dengan istilah *voice recognition*.

Azhar menyatakan bahwa pada prinsipnya bunyi ujaran terdiri atas beberapa komponen, yaitu pitch, formant dan spectrogram yang dapat digunakan untuk mengidentifikasi karakteristik seseorang untuk kepentingan *voice recognition*. Untuk kepentingan analisis

bunyi ujaran, ada beberapa komponen yang dianalisis, yaitu:

1. Pitch juga disebut dengan istilah frekuensi fundamental (dasar) dengan notasi  $f_0$ . Masing-masing orang memiliki pitch yang khas (*habitual pitch*) yang sangat dipengaruhi oleh aspek fisiologis larynx manusia. Pada kondisi pembicaraan normal, level *habitual pitch* berkisar antara 50 sampai dengan 250 Hz untuk laki-laki dan 120 sampai dengan 500 Hz untuk perempuan. Perubahan  $f_0$  secara konstan dapat memberikan informasi linguistic, contohnya membedakan antara intonasi dan emosi. Analisis pitch dapat digunakan untuk melakukan *voice recognition* terhadap suara seseorang yaitu melalui analisa statistik terhadap nilai *minimum pitch*, *maximum pitch*, dan *mean pitch*. Bunyi ujaran dapat dibedakan melalui *pitch*, kekerasan suara, dan kualitas vokal (Ladefoged dan Johnson, 2011:7). Wedhawati, dkk. (2001:24) menjelaskan bahwa kualitas bunyi vokal ditentukan oleh empat faktor, yaitu tinggi rendah lidah, bagian lidah yang bergerak, hubungan posisional

artikulator aktif dengan artikulator pasif, dan bentuk bibir. Sementara itu, Bickford dan Rick (2006:32) menyatakan bahwa kualitas vokal dipengaruhi oleh perbedaan 2 posisi lidah dengan mulut baik posisi depan-belakang maupun posisi atas-bawah. Perbedaan vokal dijelaskan oleh Ladefoged dan Johnson (2011: 22-23) dapat dilihat dari *pitch* dan nada tambahan yang berhubungan secara kasar terhadap perbedaan vokal depan dan vokal belakang. Tinggi rendah *pitch* vokal ditentukan oleh posisi lidah, *pitch* vokal tinggi adalah ketika posisi lidah rendah dan sebaliknya *pitch* vokal rendah adalah ketika posisi lidah tinggi.

2. Forman adalah frekuensi-frekuensi resonansi dari filter, yaitu *vocal tract* (*articulator*) yang meneruskan dan memfilter bunyi luaran (*output*) vokal, konsonan, atau kata. Cohn dalam Aronoff dan Janie (2003), menjelaskan bahwa bunyi vokal secara umum memiliki ciri dalam hal ketinggian lidah atau rahang (tinggi, menengah, rendah) dan bagian lidah yang bergerak (depan, tengah, belakang). Selain itu, bunyi vokal muncul karena alat ucap tidak terlalu berdekatan sehingga aliran udara tidak mengalami hambatan (Ladefoged dan Johnson, 2011). Ladefoged dan Johnson (2011) menjelaskan bahwa bunyi vokal lebih baik dijelaskan melalui penjelasan struktur akustik daripada melalui penjelasan pengaruh pergerakan secara artikulatoris.

Penelitian ini bertujuan menampilkan hasil analisis forman menggunakan analisis Praat dan membandingkannya dengan animasi *vocal tract*. Melalui perhitungan  $f_1$  dan  $f_2$  pada perangkat Praat dan animasi *vocal tract*, diharapkan dapat diperoleh nilai yang cukup presisi mengenai posisi dan pergerakan lidah dalam memproduksi ujaran sehingga dapat dilihat titik kesalahan fonetisnya.

## METODE PENELITIAN

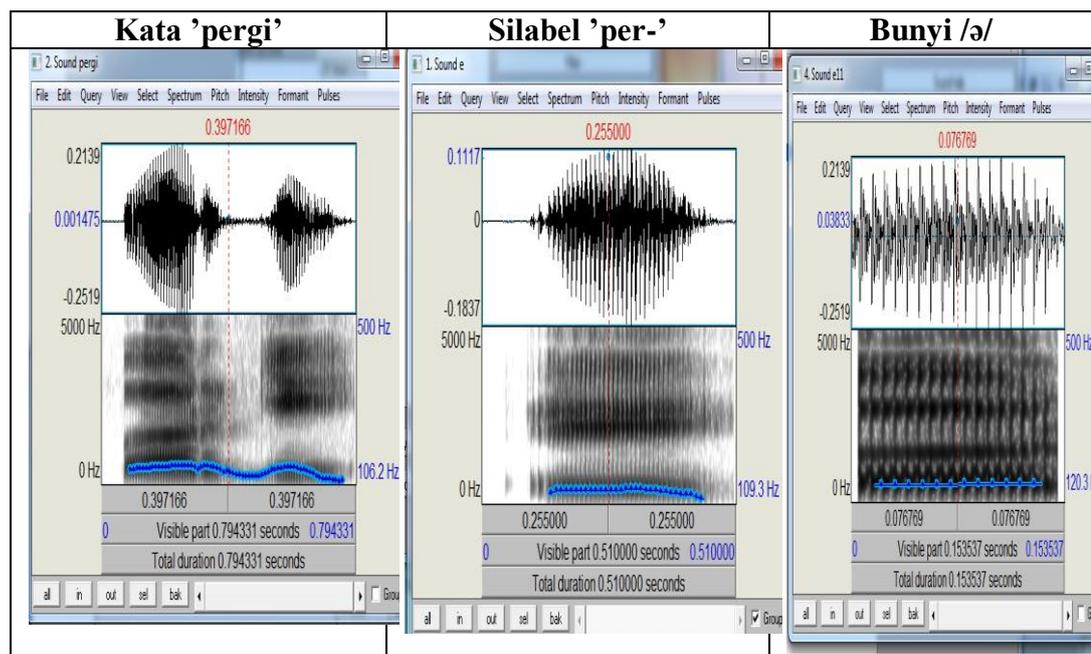
Metode penelitian yang digunakan adalah studi akustik (van Zanten dan van Heuven, 1983). Proses produksi dilakukan melalui perekaman terhadap informan native dan non native Bahasa Indonesia. Informan membunyikan vokal dan konsonan secara 2 eksperimen, yaitu membunyikan bunyi vokal tersendiri dan bunyi vokal dan konsonan berdasarkan konteks (kata tertentu). Proses perekaman dilakukan menggunakan alat perekam Sony dengan jarak perekaman ideal dan natural. Informan adalah laki-laki dewasa dan mampu membunyikan bunyi ujaran vokal dan konsonan dengan jelas. Data dianalisis menggunakan perangkat lunak Praat (analisis forman), aplikasi Visartico dan macro media flash untuk menghasilkan *vocal tract* yang telah didesain sebelumnya.

## HASIL DAN PEMBAHASAN

### Hasil penelitian

Hasil penelitian perhitungan  $f_0$ ,  $f_1$  dan  $f_2$  dilakukan dalam dua tahap. Tahap 1 adalah menggunakan perangkat lunak Praat. Selanjutnya bunyi ujaran dimasukkan dalam aplikasi Visartico untuk memperoleh dua perbandingan hasil  $f_0$ ,  $f_1$ , dan  $f_2$ .

Hasil pertama adalah analisis melalui Perangkat Lunak Praat dan telah diperoleh Contoh kurva *pitch* (lihat gambar 1) pada kata 'pergi' dibandingkan dengan kurva kontur bunyi /ə/ secara tersendiri. Gambar 1. menunjukkan hasil analisis akustik vokal /ə/ atau ê (pepet) yang diproduksi dengan durasi 0.51 dt, dengan nilai  $f_0$  106.43 hz. Kata pergi disegmentasikan dan diambil silabel per-selanjutnya disegmentasi untuk diambil bunyi /ə/. Langkah selanjutnya adalah hasil segmentasi vokal /ə/ dihitung untuk mencari nilai mean  $f_0$ , nilai min  $f_0$ , nilai max  $f_0$ . Selanjutnya dilakukan analisis forman yaitu menghitung  $f_1$  dan  $f_2$  untuk mengetahui karakteristik bunyi ujaran pada bunyi vokal secara tersendiri (diambil dari perekaman tunggal) dan bunyi vokal yang diproduksi dalam konteks (kata).



Gambar 1. Proses segmentasi bunyi /e/

Tabel 1. Nilai perbandingan Mean Pitch, nilai min, nilai max bunyi vokal (hz)

Vokal			
	mean Pitch	min pitch	max pitch
/i/(dalam konteks)	125.52	116.06	127.92
/i/ tersendiri	115.49	85.67	128.42
/a/(dalam konteks)	121.84	120.07	123.28
/a/ tersendiri	108.31	92.18	120.3
/u/ (dalam konteks)	102.89	87.26	123.97
/u/ tersendiri	114.35	85.95	124.64
/ə/ /(dalam konteks)	118.84	116.68	121.42
/ə/ tersendiri	188.69	111.76	498.7
/o/(dalam konteks)	127.73	127.33	128
/o/ tersendiri	113.41	79.33	135.6

Melalui perhitungan analisis mean pitch (min f0), nilai min f0, dan nilai max f0 pada masing-masing bunyi vokal disajikan pada tabel 1 berikut. Melalui perbandingan antara bunyi vokal tersendiri dan bunyi vokal dalam konteks diperoleh nilai mean f0 pada masing-masing bunyi vokal. Karakteristik f0 bunyi vokal /a/ dan /ə/ menunjukkan hasil perbedaan yang cukup signifikan dibandingkan dengan bunyi vokal yang

lain. Sementara perhitungan nilai f1 dan f2 dapat dilihat pada tabel 2.

Berdasarkan tabel 2 terlihat perbedaan signifikan pada produksi vokal a tersendiri dan bunyi vokal a pada konteks. Nilai yang diperoleh pada tabel di atas hampir menyerupai tabel perhitungan nilai mean f0, nilai f1, dan nilai f2 pada informan native bahasa Indonesia yang berlatar belakang bahasa Jawa. Melalui penelitiannya, Zanten mengukur

f0, f1, dan f2 secara tersendiri dan dimasukkan dengan konteks. Perolehan nilai mean f1 dan f2 pada tabel 3. Tabel 3 menunjukkan karakteristik bunyi vokal tersendiri dan bunyi vokal dalam konteks

dalam penelitian Van Zanten. Van Zanten mengkhususkan penelitiannya pada penelitian produksi vocal tract pada orang Indonesia dari berbagai kultur yang berbeda

**Tabel 2. Nilai perbandingan f1 dan f2 pada bunyi vokal**

Vokal	f1	f2
/i/(dalam konteks)	367.32	2482.93
/i/ tersendiri	426.09	2538.35
/a/(dalam konteks)	861.61	1460.681
/a/ tersendiri	120.3	1334.58
/u/ (dalam konteks)	373.21	782.85
/u/ tersendiri	389.04	74.611
/ə/ /(dalam konteks)	635.42	1411.69
/ə/ tersendiri	484.24	1353.01
/o/(dalam konteks)	511.34	1002.57
/o/ tersendiri	543.6	825.68

**Tabel 3. Perbandingan nilai bunyi vokal tersendiri dan bunyi vokal dalam konteks**

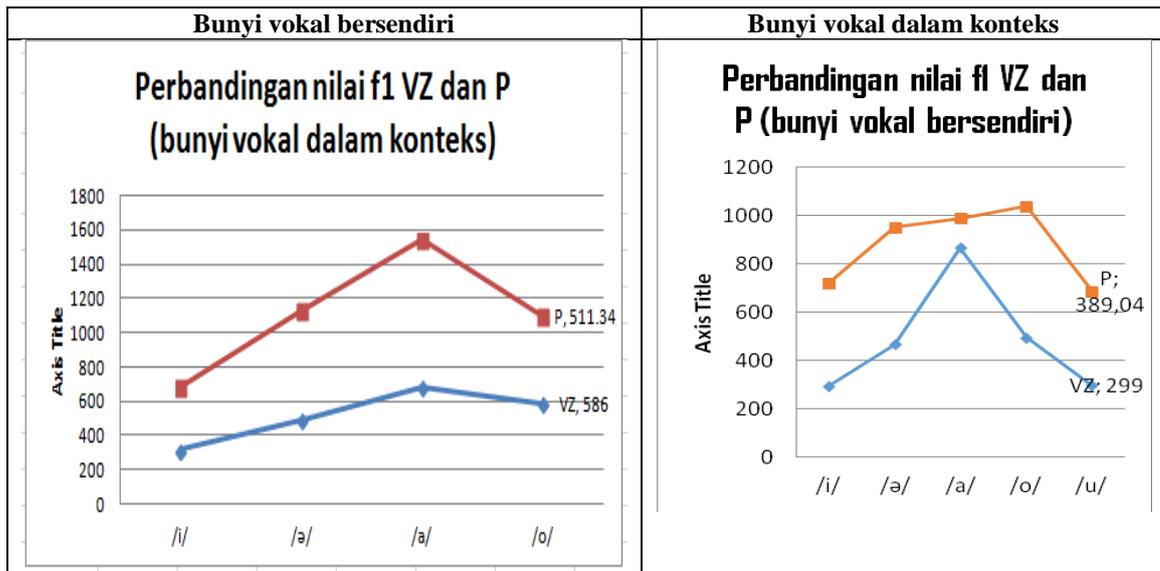
Bunyi vokal	Mean f1	Mean F2	Mean f1	Mean F2
	bersendiri	bersendiri	konteks	konteks
/i/	295	2378	316	1988
/ə/	467	1464	492	1130
/a/	868	1389	680	1142
/o/	495	1010	586	970
/u/	299	1005	342	996

**Tabel 4. Nilai f1 dan f2 tersendiri antara Van Zanten dan perhitungan peneliti**

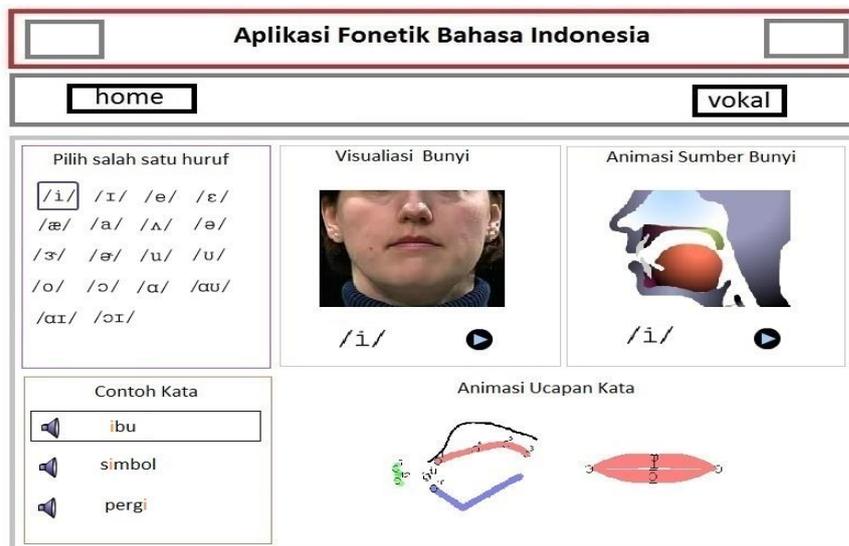
Bunyi vokal	Mean f1 (VZ)	Mean F2 (VZ)	Mean f1 (P)	Mean F2 (P)
	bersendiri	bersendiri	bersendiri	bersendiri
/i/	295	2378	426.09	2538.35
/ə/	467	1464	484.24	1353.01
/a/	868	1389	120.3	1334.58
/o/	495	1010	543.6	825.68
/u/	299	1005	389.04	74.611

**Tabel 5. Nilai f1 dan f2 dalam konteks antara Van Zanten dan perhitungan peneliti**

Bunyi vokal	VZ	VZ	P	P
	Mean f1	Mean F2	Mean f1	Mean F2
	konteks	konteks	konteks	konteks
/i/	316	1988	367.32	2482.93
/ə/	492	1130	635.42	1411.69
/a/	680	1142	861.61	1460.681
/o/	586	970	511.34	1002.57
/u/	342	996	373.21	782.85



Gambar 2. Perbandingan nilai f1 pada bunyi vokal tersendiri dan bunyi vokal dalam konteks



Gambar 3. Ilustrasi produksi vokal /a/ menggunakan aplikasi VisArtico

Tabel 4, 5, dan gambar 2 menjelaskan mengenai kecenderungan nilai yang hampir sama antara perhitungan yang dilakukan oleh Van Zanten dan perhitungan yang telah dilakukan oleh peneliti. Bunyi vokal /a/ cenderung memberikan nilai f1 dan f2 yang berbeda antara produksi bunyi vokal tersendiri atau pun produksi bunyi vokal dalam konteks. Van Zanten (1989) telah meneliti karakteristik bunyi vokal dan konsonan yang diproduksi oleh penutur dari suku Jawa, Sunda, dan Batak Toba.

Frekuensi fundamental dikenal juga dengan f0 yang koheren dalam bentuk transisi forman f1, f2, dan sebagainya.

Komponen frekuensi dominan yang mengkarakterisasi fonem-fonem yang berhubungan dengan komponen frekuensi resonansi dari sistem vokal didefinisikan sebagai forman. Suara yang terucapkan, secara khusus adalah vokal, biasanya memiliki 3 buah forman dan seringkali disebut sebagai forman kesatu, kedua, dan ketiga, dimulai dengan komponen frekuensi terendah. Ketiganya selalu dituliskan sebagai f1, f2, dan f3. forman 4 dan forman 5 dibutuhkan untuk mendapatkan nilai parameter forman yang lebih detail karena bila sinyal suara yang kita olah hanya memiliki forman yang kurang dari 3 buah, maka dapat dipastikan

analisa terhadap data tersebut akan gagal (Zanten, 1989).

Ladefoged (1975: 173) menjelaskan bahwa frekuensi formant terendah, F1, secara terbalik menggambarkan ketinggian vokal; formant kedua, F2, (atau sekedar perbedaan antara frekuensi F1 dan F2) sejalan dengan tingkat ke belakang di dalam diagram vokal yang tradisional dalam berapa hal sampai F3, sangat menentukan dalam membedakan vokal yang satu dengan yang lain. Frekuensi-frekuensi pusat ketiga diukur menggunakan spektrogram seksian pita sempit (Kay Sonograph, 6061), filter dengan lebar pita gelombang 50 Hz yang dibuat di pertengahan vokal. Pada skala yang digunakan 1 mm di sumbu frekuensi sama dengan 82 Hz.

Berikut pada gambar 3 dapat dilihat produksi vokal /a/ menggunakan aplikasi VisArtico. Melalui VisArtico letak produksi bunyi vokal lebih jelas dan mudah untuk dibayangkan. Pergerakan posisi lidah juga mendukung kejelasan posisi bunyi vokal secara presisi. Sebenarnya agak sulit menemukan ketumpangtindihan antara realisasi vokal-vokal yang diucapkan di dalam konteks, meskipun semua jenis realisasi per vokal sudah diperhitungkan. Persebaran bagi /e/, /o/, dan /u/ yang diucapkan tersendiri dan ketumpangtindihan yang lebih besar bagi vokal-vokal yang tersendiri tampak jelas, terutama pada bunyi vokal /a/ tersendiri tampak jelas, dan sebagian muncul dalam persebaran vokal lain, seperti /e/.

Melalui aplikasi VisArtico, peneliti mencoba menyajikan bentuk perhitungan F1 dan F2 dalam bentuk animasi. Dengan menggunakan contoh kata yang sama, yaitu kata 'pergi', peneliti memperoleh kecenderungan nilai yang menyerupai hasil perhitungan dari aplikasi Praat. Dalam pemerian akustik bunyi vokal terlihat frekuensi pusat formant-formant lebih rendah, yaitu kelompok overton yang berdekatan yang diperkeras oleh karakteristik resonansi saluran suara yang diukur.

## SIMPULAN

Melalui analisis produksi telah dilakukan pengukuran nilai F0, F1, dan F2 pada informan native Bahasa Indonesia. Bunyi vokal yang diukur adalah 6 bunyi vokal, yaitu /a/, /i/, /u/, /ə/, dan /o/. Bunyi vokal diukur secara tersendiri dan diukur dalam konteks yang selanjutnya dibandingkan. Hasil perbandingan menunjukkan kecenderungan nilai yang sama. Peneliti juga membandingkan pengukuran yang telah dilakukan peneliti dengan hasil pengukuran Van Zanten dengan pengukuran variabel yang sama. Secara visual, pengukuran F1 dan F2 tidak mungkin dilakukan, namun menggunakan aplikasi VisArtico dan perangkat lunak Praat, perhitungan vocal tract telah berhasil dilakukan dan telah diperoleh nilai yang presisi.

## DAFTAR PUSTAKA

- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer. (version 5.1.05).
- Clark, J., and Yallop, C. (1995). *Phonetics and Phonology*. Oxford: Blackwell. Ch 7.1-7.13.
- Denes, P.B., and Pinson, E.N. (1973). *The Speech Chain*. Murray Hill, NJ: Bell Telephone. Ch 3, 4.
- Johnson, K. (1997). *Acoustic and Auditory Phonetics*. Oxford: Blackwell. Chs. 1, 2, (3), 4, 5.
- Hayward, K. (2000). *Experimental Phonetics*. London: Longman. Chs. 2-4.
- Borden, J. & Harris, K.S. (1980) *Speech Science Primer*. Baltimore: Williams & Wilkins. Ch 3; ch 4 pp 89-130.
- Fry, D.B. (1979). *The Physics of Speech*. Cambridge: CUP. Ch 1-9.
- M. N. Al-Azhar. (2001). *Audio Forensic: Theory And Analysis*. Pusat Laboratorium Forensik Polri Bidang Fisika dan Komputer Forensik.
- Pickett, J.M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Needham

- Heights, MA: Allyn & Bacon. Ch. 2-4
- Or Pickett, J.M. (1980) *The Sounds of Speech Communication*. Baltimore: University Park Press. Ch 1-4.
- Rosen, S. and Howell, P. (1991). *Signals and Systems for Speech and Hearing*. Academic Press. Clearly written and relatively nontechnical, but much more detail than the above books. Try it if you enjoy more scientific approaches. Relevant sections are spread over several chapters—find them in Table of Contents & Index.
- Liberman, A. M. (1954). *The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants*. Washington, DC: American Psychological Association.
- Palmer, A., & Shamma, S. (2004). Physiological Representations of Speech. In S. Greenberg, W. A. Ainsworth, A. N. Popper, R. R. Fay (Eds). *Speech Processing in the Auditory System: Springer Handbook of Auditory Research* (Vol, 18, pp, xiv, 476). New York: Springer.
- Zanten, van. 1989. *Vokal-vokal Bahasa Indonesia*. Jakarta: Balai Pustaka