

KARAKTERISASI AKUSTIK TIGA BUNYI SEHARI-HARI: TEPUKAN TANGAN, UCAPAN “MULTIMEDIA”, DAN PULPEN JATUH MENGGUNAKAN AUDACITY, SONIC VISUALISER, DAN PRAAT

Acoustic Characterization of Three Everyday Sounds: Hand Claps, the Spoken Word “MULTIMEDIA”, and a Falling Pen Using Audacity, Sonic Visualiser, and Praat

Judika Marpaung

Program Studi S1 Teknik Informatika, Fakultas Teknologi Informasi, Universitas Perbanas Institute
Jakarta, Indonesia

Abstrak

Analisis karakteristik bunyi merupakan dasar pengolahan audio dalam sistem multimedia. Penelitian ini mengkarakterisasi tiga bunyi sehari-hari, yaitu tepukan tangan, ucapan kata “MULTIMEDIA”, dan pulpen yang jatuh ke meja kayu, yang direkam dengan mikrofon bawaan laptop dalam format WAV 16-bit, mono, 48 kHz. Analisis domain waktu dan domain frekuensi dilakukan menggunakan Audacity 3.7.9 dan Sonic Visualiser 5.2.1, sedangkan nada dasar dan forman ucapan diukur menggunakan Praat 7.0.02; seluruh angka diverifikasi dengan Python (metode Welch dan Parselmouth). Hasil menunjukkan bahwa tepukan tangan dan pulpen jatuh bersifat impulsif dengan waktu serangan di bawah 10 ms dan faktor puncak di atas 26 dB. Tepukan meluruh 20 dB dalam 4,5–6,5 ms dengan frekuensi dominan sekitar 1,05 kHz, sedangkan benturan pulpen memerlukan sekitar 482 ms untuk meluruh 40 dB dan memiliki spektrum paling datar serta *spectral centroid* tertinggi (2.877 Hz). Ucapan merupakan satu-satunya bunyi bernada dengan nada dasar rata-rata 113,7 Hz; puncak spektrumnya (234 Hz) adalah harmonik kedua, dan nilai F1–F2 setiap segmen sesuai dengan vokal /u/, /i/, /e/, serta peralihan /i/ ke /a/. Selisih frekuensi dominan antara Audacity dan Python paling besar 9 Hz, di bawah resolusi FFT yang digunakan. Penelitian ini juga menemukan pemotongan energi yang tajam di atas sekitar 9 kHz yang kemungkinan besar disebabkan oleh pemrosesan pada perangkat perekam.

Kata kunci: *analisis audio; spektrogram; frekuensi dasar; forman; bunyi impulsif*

Abstract

Sound characterization underlies audio processing in multimedia systems. This study characterizes three everyday sounds, namely hand claps, the spoken word “MULTIMEDIA”, and a pen falling onto a wooden desk, recorded with a built-in laptop microphone as 16-bit, mono, 48 kHz WAV files. Time- and frequency-domain analyses were performed in Audacity 3.7.9 and Sonic Visualiser 5.2.1, while the fundamental frequency and formants of the speech were measured in Praat 7.0.02; all figures were verified in Python (Welch's method and Parselmouth). The claps and the falling pen are impulsive, with attack times below 10 ms and crest factors above 26 dB. The claps decay by 20 dB within 4.5–6.5 ms with a dominant frequency of about 1.05 kHz, whereas the pen impact takes about 482 ms to decay by 40 dB and shows the flattest spectrum and the highest spectral centroid (2,877 Hz). Speech is the only pitched sound, with a mean fundamental frequency of 113.7 Hz; its spectral peak (234 Hz) is the second harmonic, and the F1–F2 values of each segment match the vowels /u/, /i/, /e/, and an /i/-to-/a/ glide. Audacity and Python differ by at most 9 Hz in dominant frequency, below the FFT resolution used. A sharp energy cutoff above about 9 kHz was also found, most likely caused by processing in the recording device.

Keywords: *audio analysis; spectrogram; fundamental frequency; formants; impulsive sounds*

1. PENDAHULUAN

Audio merupakan salah satu elemen utama sistem multimedia. Sebelum dapat diolah komputer, bunyi harus diubah menjadi data digital melalui pencuplikan (*sampling*) dan kuantisasi (Li dkk., 2021). Teorema pencuplikan menyatakan bahwa sinyal dapat direpresentasikan utuh bila frekuensi cuplik paling sedikit dua kali frekuensi tertingginya (Shannon, 1949), sehingga rekaman 48 kHz mampu memuat komponen hingga 24 kHz.

Setiap sumber bunyi memiliki ciri akustik yang berbeda. Bunyi periodik seperti vokal tersusun atas frekuensi dasar (F0) dan harmoniknya, sedangkan bunyi impulsif seperti tepukan dan benturan berlangsung singkat dengan energi tersebar pada pita frekuensi yang lebar (Rossing, 2014). Pada ucapan, resonansi saluran suara membentuk forman; dua forman pertama (F1 dan F2) menjadi penanda utama identitas vokal (Peterson & Barney, 1952). Pada tepukan tangan, Repp (1987)

menunjukkan bahwa spektrum tepukan bergantung pada konfigurasi tangan saat menepuk.

Karakteristik tersebut dapat diamati dengan perangkat lunak bebas, antara lain Audacity (Audacity Team, 2026), Sonic Visualiser (Cannam dkk., 2010), dan Praat (Boersma dkk., 2026). Namun, pengukuran dari tampilan aplikasi bergantung pada pengaturan analisis, dan kesesuaian hasil antaraplikasi jarang diperiksa secara kuantitatif dalam praktik pembelajaran. Penelitian ini bertujuan (1) mengkarakterisasi tepukan tangan, ucapan “MULTIMEDIA”, dan pulpen jatuh dalam domain waktu dan frekuensi; (2) mengukur nada dasar dan forman ucapan; serta (3) menilai kesesuaian hasil Audacity, Sonic Visualiser, dan Praat terhadap perhitungan independen dengan Python.

2. METODE PENELITIAN

2.1 Perekaman

Perekaman menggunakan laptop ASUS Vivobook 16X M1603QA dengan mikrofon bawaan (Microphone Array, Realtek Audio). Pada Realtek Audio Console, *Microphone Boost* diatur +30 dB dan *Microphone Effects* pada *Basic reception optimization*. Setiap bunyi direkam satu kali dengan Audacity 4.0.1 dan diekspor ke WAV 16-bit PCM, mono, 48 kHz: (a) tiga tepukan tangan berjeda sekitar satu detik (3,870 s); (b) ucapan “MULTIMEDIA” oleh penulis (2,300 s); dan (c) pulpen yang dijatuhkan dari ketinggian sekitar 30 cm ke meja kayu dan dibiarkan memantul hingga diam (2,800 s).

2.2 Analisis dengan Perangkat Lunak

Audacity 4.0.1 tidak menyediakan menu *Plot Spectrum*, sehingga analisis Audacity dilakukan dengan Audacity 3.7.9 pada berkas yang sama. Pengaturan analisis dirangkum pada Tabel 1. Rentang *pitch* 75–300 Hz dan batas formant 5000 Hz dipilih berdasarkan F0 terukur penulis (97–132 Hz).

Tabel 1 Pengaturan analisis

Aplikasi	Analisis	Pengaturan
Audacity 3.7.9	<i>Plot Spectrum</i>	Hann, FFT 2048 ($\Delta f \approx 23,4$ Hz), sumbu log
Sonic Visualiser 5.2.1	<i>Spectrogram</i>	Jendela 1024, <i>overlap</i> 50%
	<i>Spectrum</i>	Jendela 4096, <i>overlap</i> 75%
Praat 7.0.02	<i>Pitch</i>	<i>Filtered autocorrelation</i> , 75–300 Hz
	<i>Formants</i>	Batas 5000 Hz, 5 formant, jendela 25 ms

2.3 Ukuran Kuantitatif

Seluruh angka diverifikasi dengan Python. Level puncak dan RMS dihitung dari nilai cuplikan dalam dBFS, dan faktor puncak adalah selisih keduanya. Frekuensi dominan dan *spectral centroid* (Peeters dkk., 2011) dihitung dari spektrum rata-rata metode Welch (Welch, 1967) dengan jendela Hann (Harris, 1978) 4096 titik ($\Delta f \approx 11,7$ Hz). Waktu serangan adalah selang dari titik terakhir yang berada 20 dB di bawah puncak hingga puncak pada selubung RMS 1 ms, sedangkan peluruhan adalah selang dari puncak hingga selubung turun 20 dB (tepukan) atau 40 dB (benturan pulpen). F0 dan formant dihitung dengan Parselmouth (Jadoul dkk., 2018) (mesin Praat 6.1.38, metode *autocorrelation*) sebagai pembandingan hasil Praat 7.0.02.

3. HASIL DAN PEMBAHASAN

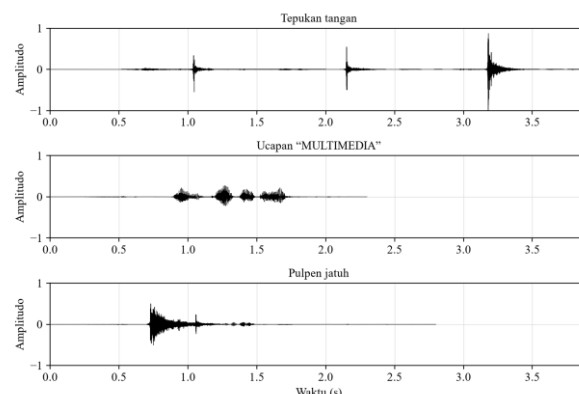
3.1 Tepukan Tangan

Ketiga tepukan muncul pada 1,046, 2,152, dan 3,179 s (Gambar 1). Setiap tepukan naik ke puncak dalam 3,0–8,0 ms dan turun 20 dB hanya dalam 4,5–6,5 ms (Tabel 2). Faktor puncak rekaman mencapai 30,6 dB. Spektrum rata-rata memuncak pada 1.052 Hz (Audacity) dan 1.043 Hz (Python), tetapi frekuensi puncak tiap tepukan berbeda-beda (261, 826, dan 914 Hz). Variasi ini sejalan dengan Repp (1987) karena konfigurasi tangan tidak pernah persis sama pada setiap tepukan. Jendela tepukan

memuat sekitar 98,9% energi rekaman, sehingga derau latar di antara tepukan tidak memengaruhi frekuensi dominan.

Tabel 2 Karakteristik setiap tepukan

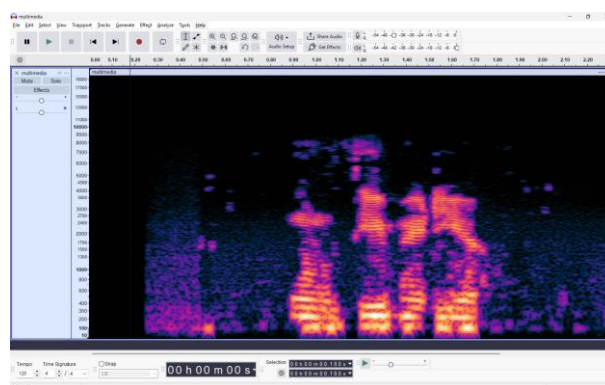
Tepukan	Serangan (ms)	Turun 20 dB (ms)	Puncak (Hz)	Centroid (Hz)
1	8,0	4,5	261	798
2	4,5	6,5	826	1.763
3	3,0	6,0	914	1.013



Gambar 1 Bentuk gelombang ketiga bunyi
Sumber: Pengolahan penulis

3.2 Ucapan “MULTIMEDIA”

Ucapan berlangsung sekitar 0,9–1,7 s dan terbagi menjadi empat kelompok suku kata (MUL, TI, ME, DIA) dengan puncak –11,3 dBFS dan faktor puncak 20,1 dB. Spektrogram (Gambar 2) memperlihatkan garis-garis harmonik yang sejajar di setiap suku kata. Nada dasar rata-rata 113,7 Hz (median 112,7 Hz; rentang 96,8–131,9 Hz), dan puncak spektrum bagian bersuara terletak pada sekitar 111, 240, 363, dan 463 Hz, mendekati kelipatan 115 Hz. Dengan demikian, frekuensi dominan 234 Hz pada Audacity adalah harmonik kedua, bukan nada dasarnya.



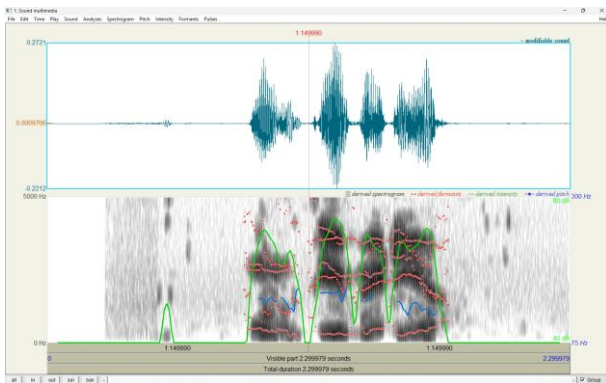
Gambar 2 Spektrogram ucapan “MULTIMEDIA” pada Audacity
Sumber: Tangkapan layar Audacity 3.7.9

Nilai formant (Tabel 3) sesuai dengan pola vokal menurut Peterson dan Barney (1952): MUL memiliki F1 dan F2 rendah (/u/), TI memiliki F1 terendah dan F2 tinggi (/i/), dan ME berada di antaranya (/e/). Pada DIA, F1 naik dari 247 ke 508 Hz dan F2 turun dari 2.206 ke 1.652 Hz, menandakan peralihan /i/ ke /a/. Suku kata TI

memiliki F0 (122,3 Hz) dan intensitas tertinggi, yang mengindikasikan penekanan. Kurva Praat (Gambar 3) konsisten dengan hasil ini.

Tabel 3 Segmen bersuara ucapan

Segmen	Waktu (s)	F0 (Hz)	F1 (Hz)	F2 (Hz)
MUL /u/	0,90–1,08	111,8	412	1.053
TI /i/	1,19–1,32	122,3	296	2.228
ME /e/	1,37–1,47	113,3	389	1.901
DIA /i→a/	1,54–1,70	108,9	346	2.248



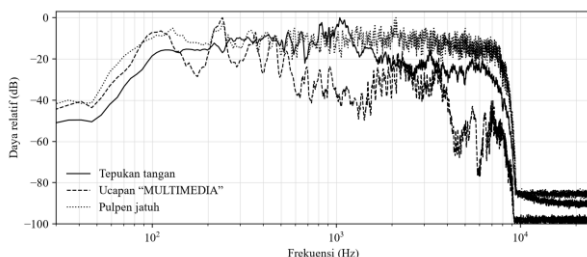
Gambar 3 Pitch, intensitas, dan forman pada Praat
Sumber: Tangkapan layar Praat 7.0.02

3.3 Pulpen Jatuh

Benturan pertama terjadi pada 0,733 s dengan serangan 9,5 ms, diikuti benturan kecil berulang (0,81–0,94 s) dan pantulan yang jelas pada 1,058 s, atau 0,325 s setelah benturan pertama. Benturan memerlukan sekitar 482 ms untuk meluruh 40 dB, jauh lebih lama daripada tepukan, karena badan pulpen dan meja kayu ikut bergetar (Rossing, 2014). Frekuensi dominannya 2.104 Hz (Audacity) dan 2.109 Hz (Python), dengan *spectral centroid* 2.877 Hz yang tertinggi di antara ketiga bunyi. Bunyi lemah periodik pada 1,27–1,47 s (± 127 dan ± 235 Hz) tidak dapat dipastikan sumbernya dari rekaman.

3.4 Perbandingan dan Kesesuaian Alat

Gambar 4 menunjukkan bahwa ucapan memiliki deretan puncak harmonik di bawah 1 kHz, tepukan memusatkan energi di sekitar 1 kHz, dan pulpen memiliki spektrum paling rata hingga sekitar 8 kHz. Tabel 4 merangkum angka kunci. Selisih frekuensi dominan antara Audacity dan Python paling besar 9 Hz, lebih kecil daripada jarak antartitik FFT 2048 pada Audacity (23,4 Hz), sehingga kedua hasil konsisten.



Gambar 4 Perbandingan spektrum rata-rata (Welch, jendela 8192)

Sumber: Pengolahan penulis

Tabel 4 Ringkasan dan kesesuaian alat

Besaran	Tepukan	Ucapan	Pulpen
Puncak (dBFS)	−0,07	−11,30	−5,87
Faktor puncak (dB)	30,6	20,1	26,3
Dominan, Audacity (Hz)	1.052	234	2.104
Dominan, Python (Hz)	1.043,0	234,4	2.109,4
Selisih (Hz)	9,0	0,4	5,4
Centroid (Hz)	1.267	497	2.877
F0 (Hz)	—	113,7	—

3.5 Keterbatasan

Pada tepukan dan benturan pulpen, energi anjlok tajam di atas sekitar 9 kHz. Pada pulpen, level turun sekitar 68 dB dari pita 8–9 kHz ke 10–12 kHz. Karena terjadi pada dua bunyi yang sangat berbeda, penyebab yang paling mungkin adalah mikrofon laptop atau pemrosesan *driver* Realtek, bukan sumber bunyinya. Selain itu, setiap bunyi hanya direkam satu kali dan nilai desibel bersifat relatif karena mikrofon tidak dikalibrasi.

4. KESIMPULAN

Tepukan tangan dan pulpen jatuh merupakan bunyi impulsif berpita lebar, tetapi berbeda pada peluruhan dan sebaran frekuensinya: tepukan hilang dalam beberapa milidetik dengan energi terpusat di sekitar 1 kHz, sedangkan benturan pulpen bergema ratusan milidetik dengan spektrum paling datar dan paling “terang”. Ucapan “MULTIMEDIA” adalah satu-satunya bunyi bernada (F0 113,7 Hz) dengan forman yang sesuai vokal penyusunnya. Audacity, Sonic Visualiser, dan Praat memberikan hasil yang konsisten dengan perhitungan Python dalam batas resolusi FFT. Penelitian lanjutan disarankan merekam setiap bunyi berulang kali dengan mikrofon eksternal tanpa pemrosesan *driver* agar komponen di atas 9 kHz dapat dianalisis.

DAFTAR PUSTAKA

- Audacity Team. (2026). *Audacity* (Versi 3.7.9) [Program komputer]. <https://www.audacityteam.org/>
- Boersma, P., Weenink, D., & Schupak, A. (2026). *Praat: Doing phonetics by computer* (Versi 7.0.02) [Program komputer]. <https://praat.org>
- Cannam, C., Landone, C., & Sandler, M. (2010). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. Dalam *Proceedings of the 18th ACM International Conference on Multimedia* (hlm. 1467–1468). ACM. <https://doi.org/10.1145/1873951.1874248>
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1), 51–83. <https://doi.org/10.1109/PROC.1978.10837>
- Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15. <https://doi.org/10.1016/j.wocn.2018.07.001>
- Li, Z.-N., Drew, M. S., & Liu, J. (2021). *Fundamentals of multimedia* (Edisi ke-3). Springer. <https://doi.org/10.1007/978-3-030-62124-7>

- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902–2916. <https://doi.org/10.1121/1.3642604>
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *The Journal of the Acoustical Society of America*, 81(4), 1100–1109. <https://doi.org/10.1121/1.394630>
- Rossing, T. D. (Ed.). (2014). *Springer handbook of acoustics* (Edisi ke-2). Springer. <https://doi.org/10.1007/978-1-4939-0755-7>
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10–21. <https://doi.org/10.1109/JRPROC.1949.232969>
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73. <https://doi.org/10.1109/TAU.1967.1161901>