International Journal of Latest Research in Humanities and Social Science (IJLRHSS)
Volume 08 - Issue 11, 2025
www.ijlrhss.com || PP. 39-57

Acoustic Tendencies through Praat

Eri Kondo[1], Kamran Mohammad[1], Mohsin Ali[2], Takehito Hayami[1]

[1]*Graduate School of Environmental Engineering, the University of Kitakyushu*

[2]*Graduate School, The Kyushu Institute of Technology*

Abstract: This report presents a trial to visualize acoustic features of particular recorded speech sounds including the nasal consonant,*n*, in a classroom by using Praat, a software developed to analyze phonetic feature in a voice waveform. The waveforms were obtained in an English class of a high school. In a textbook, three words including /n/ (monkey /ˈmʌŋ·ki/, kangaroo /kæŋ·ɡəˈru/, Jane /ˈdʒeɪn/) were typed with illustration. The students uttered the words, and the voice was recorded through the tablet devices. Using the Praat, the perspectives of time, frequency, and spectrum of the words uttered by the students were compared. We considered that it would be useful for pronunciation instruction as it allows for visual confirmation of syllable pronunciation.

1. Introduction

This paper introduces that acoustic features should be analyzed and argued on basic values on the time, frequency, and spectrum from a point of view of signals, with an experiment of recording audios of words and sentences on a tablet. The applied acoustic elements have severely transformed to visual waves and the step by step analyses through functions of Praat. Then, a feature of nasal consonant, /ŋ/ at the final position of the first syllable in CVC (consonant-vowel-consonant) structure was illustrated.

Historically, acoustic experiments have been performed on many laboratories for measuring frequency and the tendency. Two experiments and procedures are clarified with the results. First, a study this paper indicates for assessing each duration and intensity was conducted using a word list of 75 tokens containing specific words, with each word spoken five times by two male native Taiwanese speakers [1]. These recordings were made in a sound-treated room and later analyzed using a PDP-12 computer at the UCLA Phonetics Laboratory, where fundamental frequency and RMS amplitude were measured every 10 ms using the Cepstrum method as the user. To compare varying utterance lengths, the measurements were organized into five sections, each with approximately equal numbers of 10 ms intervals. The means of these intervals were calculated to determine the average fundamental frequency and RMS amplitude for each section, based on the 15 repetitions of each syllable. The experiment aimed to investigate the correlation between fundamental frequency, intensity, and duration using Taiwanese, a tone language. Second, listeners participated in an experiment, where they were exposed to 4-formant /hVd/ syllables in both unmodified and whispered conditions to explore the utilization of spectral information without f0 information [2]. The experiment aims to assess how social knowledge and gender stereotypes influence the use of spectral cues. Additionally, it evaluates the stability of speaker height judgments influenced by random between-speaker variation, with listeners divided into groups based on speaker accuracy (high versus low). It is anticipated that the high-accuracy group will produce more accurate height estimations, while both groups will exhibit similar use of acoustic information. Without doubt, recording experiments on acoustics need a tool that gets such scales and physic values. This report explores the issue of Praat is made by two phonetic scientists in Amsterdam for an applied speech perceptible analysis on pitch, intensity, formants, and spectrums. Then, Table 1 illustrates that there are advantages on optical and valuable analyses. Users who analyze a speech production need to a spectrum as a vital analysis and acoustic focuses on Praat [3]. It enables us to observe speech processing signals on pitch and formants. The dynamic way after recording utterances is often used as a direct voice method; however, it is also practical to obtain audios from remote PC and Phone [4].

Table 1: The advantages of functions on usages of Praat

Analysis	Visual	Auditory
Duration	Distance	Time
Spectrogram	2D plot	Tone quality
Pitch	Numerical value	Tone pitch
Intensity	Numerical value	Sound pressure
Formants	1D Line	Phoneme

Mainly, this report focuses on an identification of nasal consonant that is preceded by vowels of varying qualities. It highlights a previous trial on nasal consonants' perception of categories of /m/, /n/, and /ŋ/ at a final position of syllable indicating that each formant transition is crucial for identifying nasal consonants, regardless of whether it occurs before or after the consonants, and whether the stimuli are synthetic or natural speech sounds[5]. Identification and discrimination of synthesized nasal consonants by adult American subjects revealed that categorical perception occurs based on place-of-articulation contrasts, indicated by sharp boundaries in identification functions. Then, the experiments demonstrated that the discrimination accuracy was lower for unfamiliar syllable-initial contrasts compared to familiar syllable-final contrasts, while no difference was noted for oral stop consonants. This suggests that familiarity with specific phonetic contrasts influences adults' perception of relevant acoustic dimensions in different phonological contexts. Then, the issue has become more important as syllable structures are recognized as a contrast argument on an acoustic experiment. An approach is completely identified with the idea that stressed and unstressed positions have an effect on rhythm patterns, related to acoustic perception and production from a movement of articulation of jaw [6]. Therefore, the movement of jaw is important for the articulation. A pitch and duration of vowels are mainly based on stressed syllables. In addition, a vowel duration is investigated with affection the vibration of the following consonant, on /apt/, /apd/, and /abt/ [7]. On the other hand, the status of the voicing postvocalic stop consonants has an impact on the preceding vowel as /kod/, /kot/, /kog/ [8]. Moreover, a vowel quality has a link with nasal consonants on a final position of /m/ and /n/, where the vowels as /o/ and /u/ have a distinctive characteristic from the vowels as /i/ and /e/ [9]. In addition, a voice onset time in Spanish speakers show a contrasting quality between consonants of voiced /bdg/ from voiceless consonants /ptk/ in word-initial position [10]. The important factor of phonetic syllable structure is relative to a quality of schwa before a nasal, /n/ like*sudden, channel, motion,* and *uncle* [11]. Then, a vowel in the syllable does not exist and it is named 'syllabic consonants' that the articulation of nasal controls the syllable speed. Although there is some truth to these ideas, it is also true to say an articulation of vowels and consonants should be discussed with syllables.

In words, phonemes' replacements from four locations were analyzed for their shared contexts [12]. For instance, in "spar" and "saw," the phonemes /p/ and /ə/ share a common post-initial context preceded by /s/, while the phonemes, /p/ and /ɔː/ in "pray" and "oral" share a common initial context followed by /r/. The analysis quantified the common contexts between each pair of phonemes across the specified locations. The findings suggest that an initial phoneme inventory for English can be established through a method not based on vowel, consonant, or syllable considerations. By analyzing the combinatory potential of phonemes, two classes emerge that align with the traditional vowel-consonant dichotomy. The study identifies a procedure for creating a set of freely combining units, both simple and compound. It defines the syllable as the smallest combination pattern with a vowel as the nucleus, surrounded by consonants. Longer sequences are treated as a series of syllables, using the frequency of consonant combinations to ascertain syllable boundaries, independent of obvious accents. The conclusion that consonants and vowels are represented independently does not discount the importance of sonority in speech production[13]. Sonority influences consonant ordering in syllables and their boundaries, and it is linked to speech error patterns in aphasia. This indicates that the sonority and consonant/vowel structure are utilized at different levels in speech production. Additionally, the distinction between consonants and vowels is essential for prosodic structure and syllable organization. Substantial research explores the impact of lexical stress and tones on a word recognition and a speech segmentation. A speech segmentation involves defining boundaries between linguistic units in spoken language, which lacks clear separations, leading to ambiguities such as in the phrases "a nice bag" and "an ice bag" that can sound identical [14].

Drawing and getting acoustic signals for revealing a connection of structuring units, we will recognize a visualization of speech notes that there may be a distinction between consonants and vowels, related with stressed syllables and sentence intonations, or not. For right speech communication, the scientific fields should be more boarded in the future academics. The issue brings us a view that we need a tool that can measure speech productions directly.

This short experiment shows points out acoustic keys, in which a syllabic problematic combination with an articulation of consonants and vowels is recognized, in particular, in nasals.

2. Method

2.1 Procedure

- High school students (just girls in a nurse class) in Kindai Fukuoka high school participated in recording each audio through tablet's recording and listening functions. The textbook and listening audios were scanned and used at the chapter of Zoo. Words of *monkey* (mʌŋ·ki) and *kangaroo* (kæŋ·ɡəˈru), and a sentence of *Look! It's a kangaroo* were recorded. The phonetic features were not shown to students, then the listening materials were given to them.

- Each audio file was downloaded to wave. And analyzed from views of spectrum, formants, and voice onset time, speed distribution.
- High school students in Tokai Fukuoka high school participated in recording a sentence of *I'm Jane* from listening an audio of a text book.

2.2 Tablet

A tablet is used as submitting each assignment, as seen in Figure 1. The functions with recording and listening to audios have been already widespread. A tool to record voice is quite convenient; however, a tablet is useful for taking a clear speech production without noises.

3. Results

3.1 Voice onset time

The scores indicate that each measured point as voice onset time on Praat shows there is an average time as mean in Table 2. In addition, it refers to each time pace that the graph in Figure 1 represents the time characteristics of each sound rising and falling together in utterances of *monkey* and *kangaroo*. The vertical axis is time and the horizontal axis is the individual. The numbers fluctuated wildly. That is, it takes about 1 second for a student in No. 15 to utter each word; however, it takes 0.6 second for a student in No. 23 to do that. Therefore, it is not based on the vowel and consonant acoustic features although in general the vowel has a long acoustic signal. On the other hand, students in No. 1 and No. 11 are counted that it takes shorter time to utter the word, *monkey*. Then, they must have just recognized that acoustic features on each word were existing. Later, we'll pick up each student's waveform and spectrum feature which gives us their phonetic tendency.

Table 2: Voice onset time of words and sentences

	N	Mean	Std. Deviation	Std. Error Mean
monkey-word	23	.67467387	.107175941	.022347729
kangaroo-word	23	.72976183	.107457872	.022406515

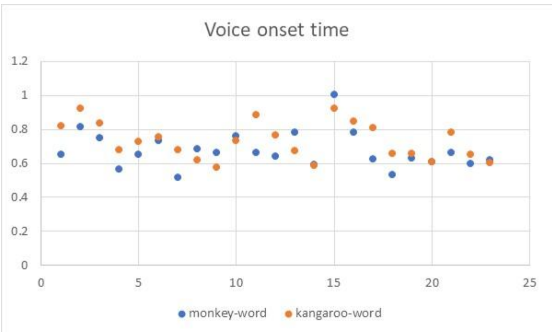

Figure 1: Voice onset time of *monkey* and *kangaroo* for each subj.

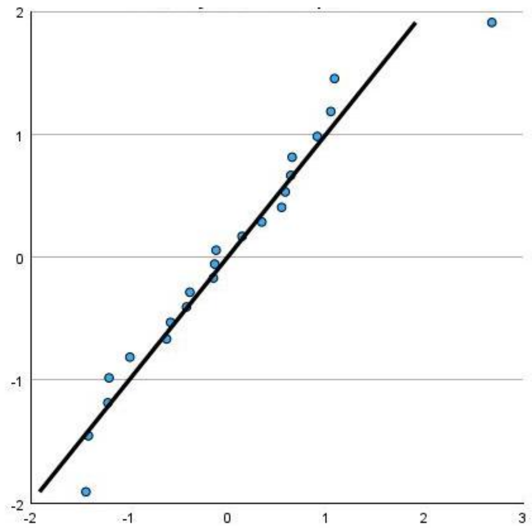

Figure 2: QQ plot of *monkey*

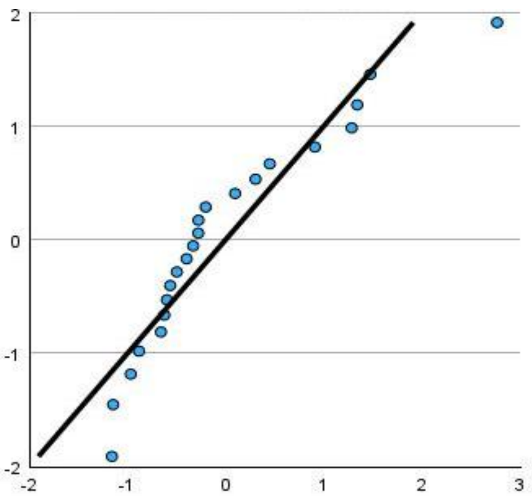

Figure 3: QQ plot of *kangaroo*

Figure 2 and 3 provide information about a distribution from an average. The reason why the time is measured is that this experiment wants to indicate that human utterances have a distribution and a difference from the average line. Figure 3 indicates that there are students who are far from the average. This experiment on a tablet's function was dependent on a listening material of audios; however, it did not have an articulation lesson for a focus on phonetic structures. With it, it would relatively illustrate data draws a technical learning change after acquiring the articulation as a learner.

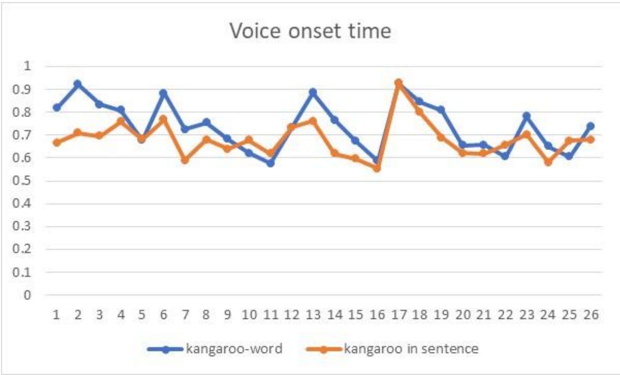

Figure 4: Voice onset time of *kangaroo* for each subj.

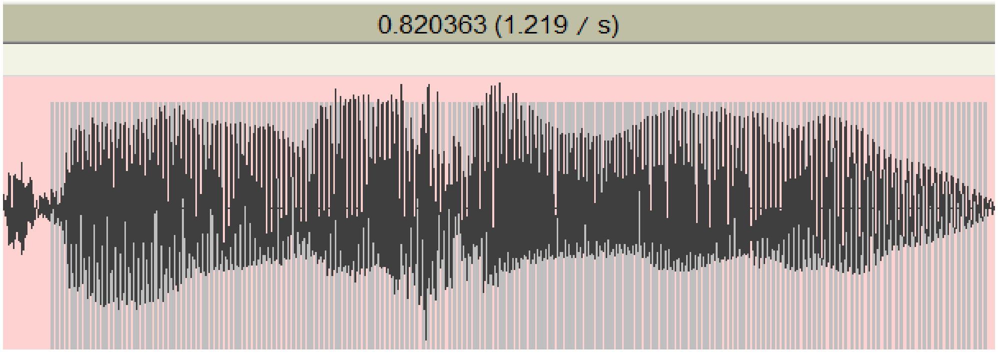

Figure 5: A student of No. 1 of *kangaroo*

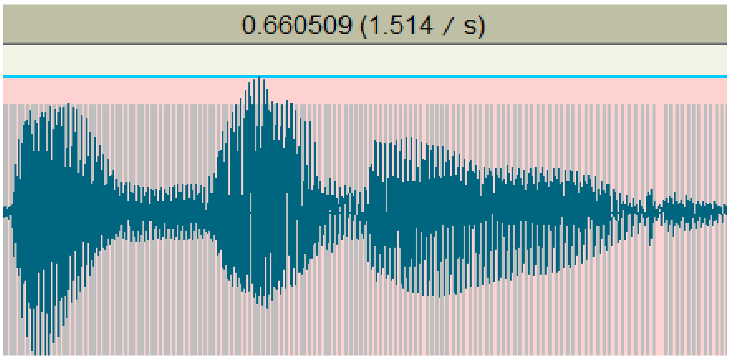

Figure 6: A student of No. 7 of *kangaroo*

As the result of recording a word of *kangaroo* and a sentence of *It's a kangaroo*, in Figure 4, where the vertical axis is time and the horizontal axis is the individual., students of No. 5, No. 11 and No. 16 reached around 0.6 second; however, No. 1 and No. 7 obviously have differing attitudes about each utterance on Figure 5 and 6. The phonetic feature of /ru/ which is contrasting with /kæŋ/ and /gə/. Moreover, the first stressed syllable is accented in Figure 6. Figure 5 indicates that the diagram shows syllables connected by articulation of vowels. Therefore, looking at the feature on voice onset time, we often consider of each characteristic on wave form and spectrum that should be recognized as the syllabic structure and articulations of vowels and consonants.

3.2 Waveform's views

The usage of waveform is quite supportable. The wide waveform is identified with a vowel feature. These waveform viewpoints display significant features from Figure 7 to 10 and Figure 11 to 14: (1) a pause between a first syllable, /mʌŋ/ and a second syllable, /ki/, and (2) an acoustic connection between /kæŋ/ and ·/gəˈru/ at the relationship of CC, /ng/.

(a) The spectrum results represent that the stressed syllable of /mʌŋ/ has been recognized as one unit to the following the syllable of /ki/.

(b) The two syllables are found at /mʌŋ/ and /ki/. Then, the same level is appeared.

(c) The pause between two syllables is noted.

(d) On the other hand, as a consequence of recording of *kangaroo*, the waveform illustrates that there is no division between the first unstressed and the second stressed syllables, and they are joined in a straight line.

(e) The nasal consonant in final syllable position has a different viewpoint, where the vowel positions as the articulation of tongue and mouth create the shift.

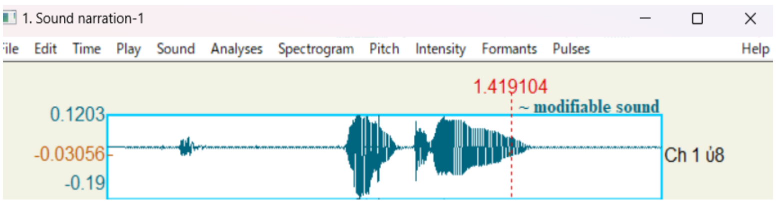

Figure 7: A waveform of *monkey* (1)

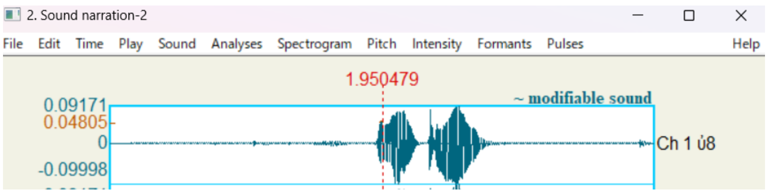

Figure 8: A waveform of *monkey* (2)

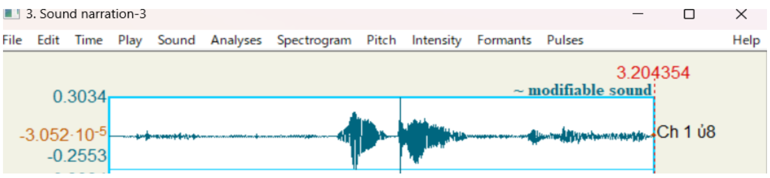

Figure 9: A waveform of *monkey* (3)

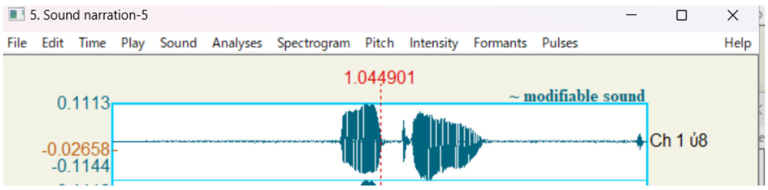

Figure 10: A waveform of *monkey* (4)

Waveforms in Figure 7 and 10 illustrate that a strong long speech feature appears at both the first and second syllables and there is a space of pause after the first syllable, *mon*. In addition, Figure 8 and 9 draw a shorter illustration although they have a pause there.

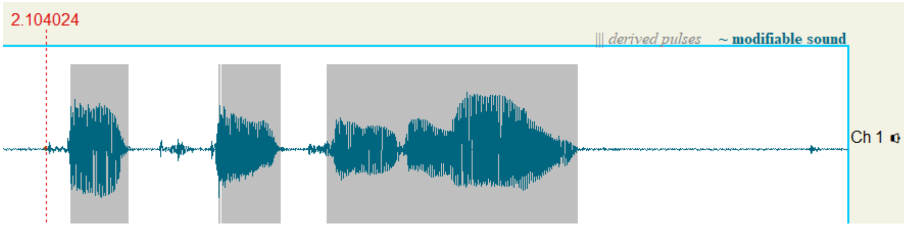

Figure 11: A waveform of *kangaroo* (1)

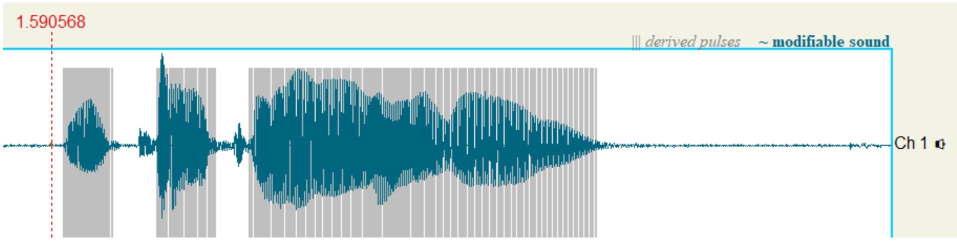

Figure 12: A waveform of *kangaroo* (2)

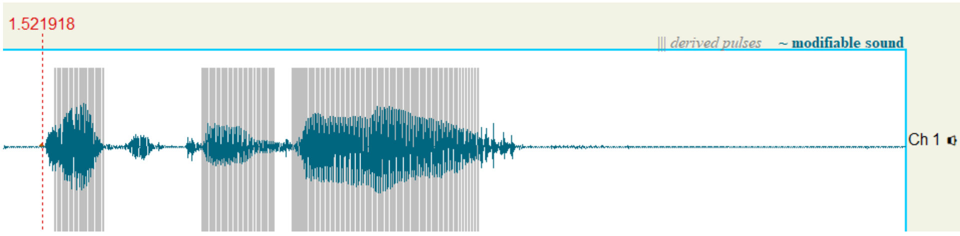

Figure 13: A waveform of *kangaroo* (3)

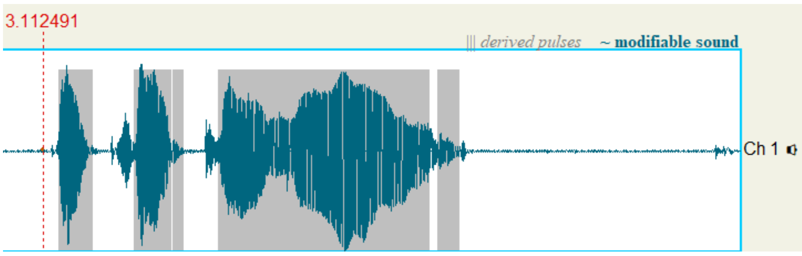

Figure 14: A waveform of *kangaroo* (4)

Waveforms in Figure 12 and 13 have a similar image of misunderstanding syllable structures of /kæŋ·gəˈru/ to /kæŋgəˈ·ru/, then /ru/ has a strong feature. Waveforms in Figure 11 and 14 indicate that the first syllable /kæŋ/ is recognized: however, the vowel of /u/ sound in the second syllable has a strength.

This paper represents that words with two syllable structures containing /n/, with a nasal alveolar feature as an articulation position at the middle place of words have a different acoustic feature, in the line with the following plosive velar consonants, /k/ (voiceless) and /g/ (voiced) on each relation of CC of *nk* and *ng*. They have stressed and unstressed syllabic distinctions between them, where the first stressed syllable /mʌŋ/ contains the consonant, /ŋ/ at the final position, on the other hand, the second stressed syllable contains the consonant, /g/ on /kæŋ·gəˈru/. The CVC structures of /mʌŋ/ and /kæŋ/ in words have the different distinction as one stressed syllable inside two syllables and a connection to each next consonant, /k/ and /g/. Then, the consonants of /k/ and /g/ is belonging to the second syllable that connects to the consonant, /ŋ/. Then the CVC/CV structures having /ŋk/ and /ŋg/ indicate that there is a pause between /ŋ/ and /k/ and the CC connections of /ŋg/ do not have such a pause. The latter is related with the nasal feature, /ŋg/. It is also based on the articulation's feature of the following vowels, /i/ and /ə/, related to the highest position of tongue of /ə/. Therefore, the tendency from spectrum's results indicate that the articulation and the position of tongue are manipulated at a different way, and the articulation of consonants and vowels influence each other and manifest as acoustic trends. The distinguishable waveform displays often give us a significant distinction as a vision of specifying a sound wave. Moreover, the result also indicates that the preceding vowel's feature often has the syllable structure on a difference between /mʌŋ/ and /kæŋ/. In addition, the following vowel feature of plosive consonants /k/ and /g/ has also an important role on acoustic characteristics on /i/ and /ə/ because of the back position of tongue in a mouth.

3.3 Spectrum's views

Spectrum views paint that the vocal intensity of has a dark side. In addition, the frequency is gained from a button of a view spectrum slice. This time, looking at the spectrum shots of students of No. 15 who was speaking slowly and of No. 1 who was speaking at the average time, we compare the former intensity with the latter. Figure 15 illustrates that the student has a strong intensity on vocal tract, compared with Figure 17. The intensity can draw a spectral slice which gives us a frequency of formants in Figure 16. The measurement of frequency at intensity is decreasing.

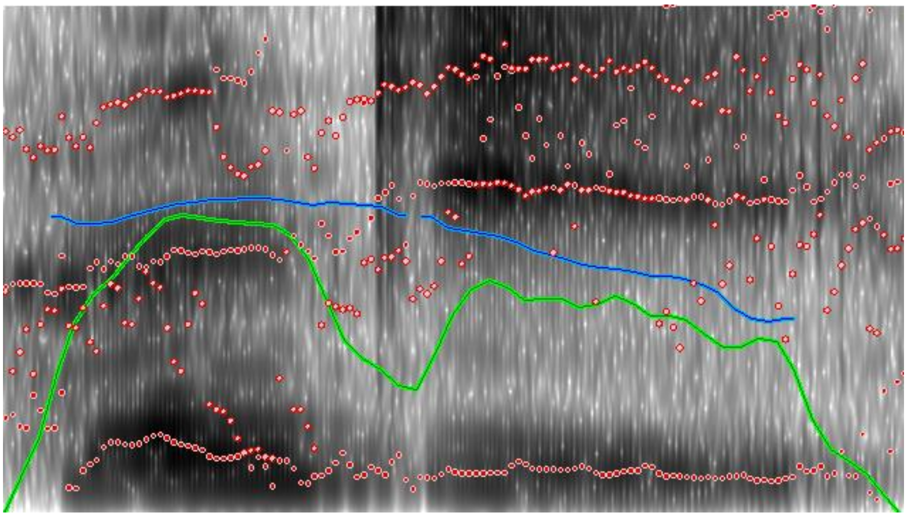

Figure 15: A spectrum of *monkey* of No. 15

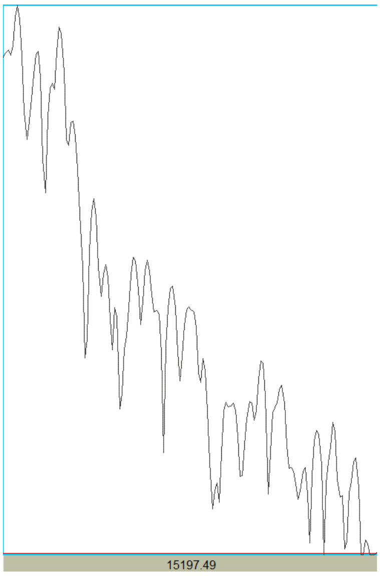

Figure 16: A view spectral slice of *monkey* of No. 15

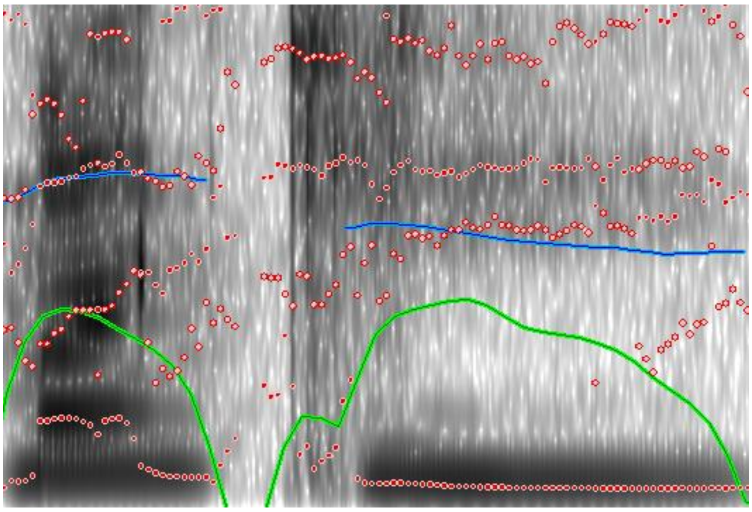

Figure 17: A spectrum of *monkey* of No. 1

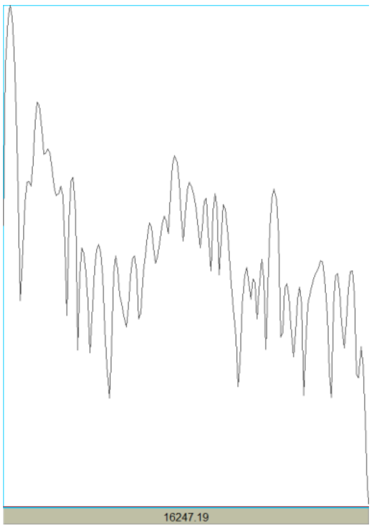

Figure 18: A view spectral slice of *monkey* of No. 1

Moreover, looking at Figure 19 and 20, we found that there is a distinctive darkness at the left side on vocal tract a student who uttered the word slowly in Figure 19. The darkness sides themselves have just given us a tendency that the speech production has been strengthen during recording the speech. Compared to Figure 16, Figure 18 is not declining, but is following a parallel line. The dark side has a characteristic of vocal tract, then it draws that there is a different dark place.

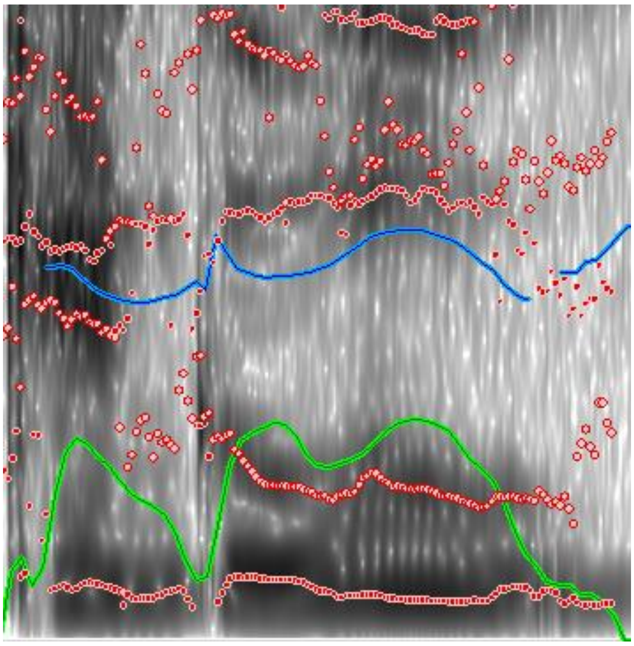

Figure 19: A spectrum of *kangaroo* of No. 2

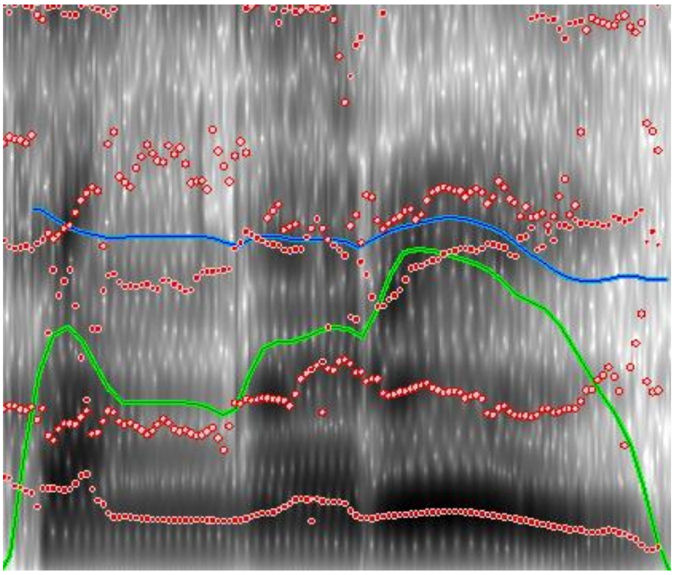

Figure 20: A spectrum of *kangaroo* of No. 14

3.4 Formant's views

Formants are valuable when we measure the highest and lowest frequency on speech. F2 and F3 are intermediate positions. It is based on the articulation and harmony. The quality of speech indicates that there is an individual difference on producing sounds that bring us an applied acoustic frequency. The frequency is said to a reference of having a low frequency in bilabial and a demonstration of having a relatively higher frequency in alveolar.

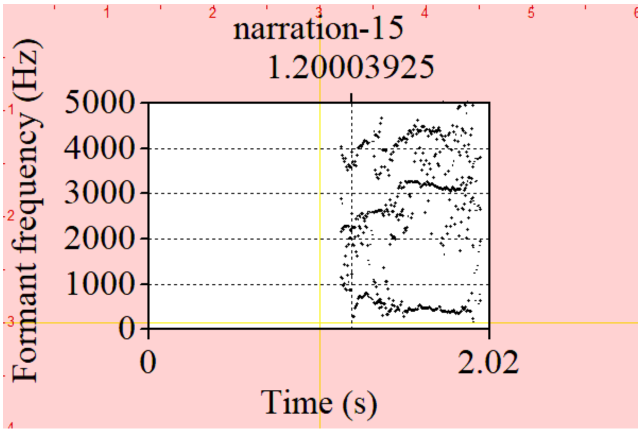

Figure 21: Formants of *monkey* (a student, No. 15)

(Time_s 1.200039 F1_Hz 850.742849 F2_Hz 1998.835964 F3_Hz 2162.383844 F4_Hz 3568.465782)

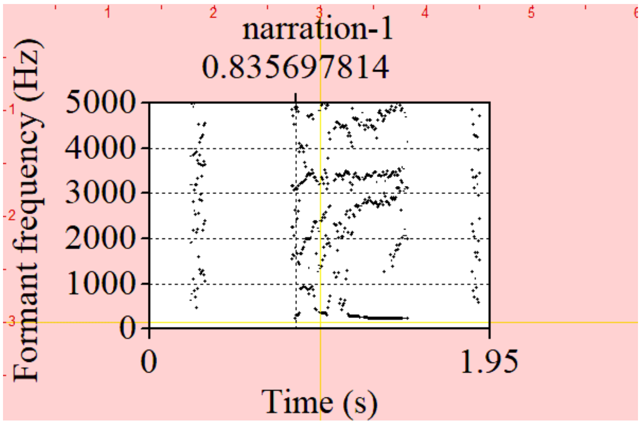

Figure 22: Formants of *monkey* (a student, No. 1)

(Time_s0.828386 F1_Hz 1608.874450 F2_Hz 2129.525982 F3_Hz 3100.149149 F4_Hz 4594.828464)

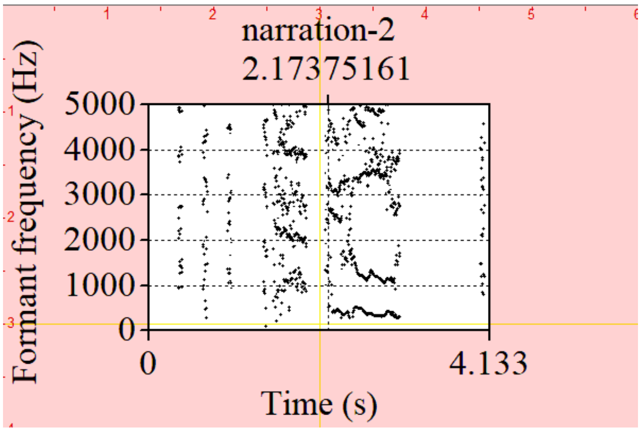

Figure 23: Formants of *kangaroo* (a student, No. 2)

(Time_s2.173752 F1_Hz 1284.984477 F2_Hz 2419.454024 F3_Hz 3133.832966 F4_Hz 4109.785518)

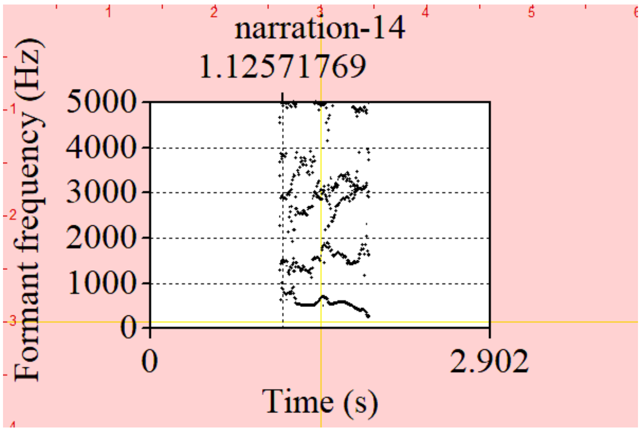

Figure 24: Formants of *kangaroo* (a student, No. 14)

(Time_s1.125718 F1_Hz 950.666527 F2_Hz 1469.419136 F3_Hz 2913.264624 F4_Hz 3739.922378)

In Figure 21 and 22 the values of Hz can be measured at each Hz. F1 and F3 show each line as frequency. Figure 23, 24 and 25 draw a difference in F2' value. In addition, Figure 26 shows that there is a difference in F1 and F2.

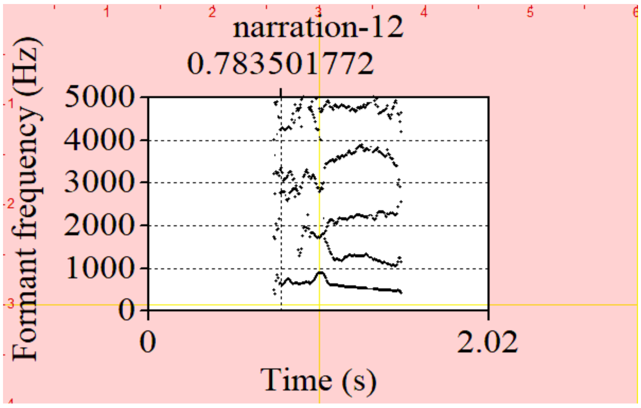

Figure 25: Formants of *kangaroo* (a student, No. 12)

(Time_s0.788551 F1_Hz 639.060710 F2_Hz 2746.082689 F3_Hz 3292.853782 F4_Hz 4233.648944)

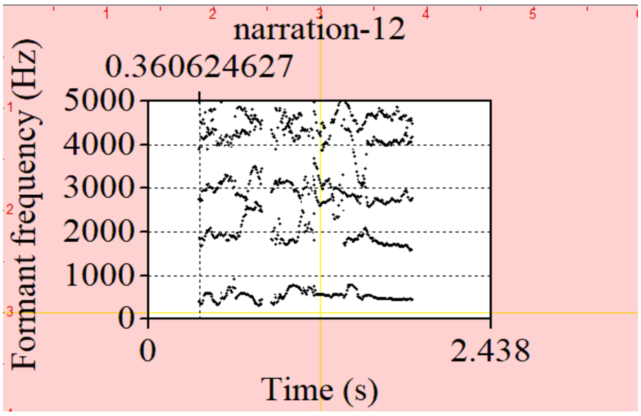

Figure 26: Formants of *It's a kangaroo* (a student, No. 12)

(Time_s0.360625 F1_Hz 457.988462 F2_Hz 1806.271071 F3_Hz 2729.943821 F4_Hz 3942.845575)

Table 3: Formants of *monkey*

	N	Mean	Std. Deviation	Std. Error Mean
F1mon	5	606.0358	444.80607	198.92332
F2mon	5	1818.4209	441.02894	197.23414
F3mon	5	2842.6037	344.79300	154.19612
F4mon	5	3911.0335	416.98524	186.48147

Table 4: Formants of *kangaroo*

	N	Mean	Std. Deviation	Std. Error Mean
F1	5	918.6746	428.02423	191.41825
F2	5	2128.4071	487.82684	218.16279
F3	5	2920.0210	183.98274	82.27958
F4	5	4009.9707	171.23994	76.58083

Table 3 and 4 indicate that five students' formants have mean and SD. The measured formants demonstrate that there is a slight difference in SD of F3 and F4 in table 4. The vocal vibration judges that the formants are dependent on the articulation's distinction.

3.5 Pitch analyses

This section explains that two features.

(1) Figure 28 to 30 show each pitch of a sentence, *It's a kangaroo*, from three students.

(2) Figure 31 to 33 indicate each pitch of a sentence, *I'm Jane*, from three students.

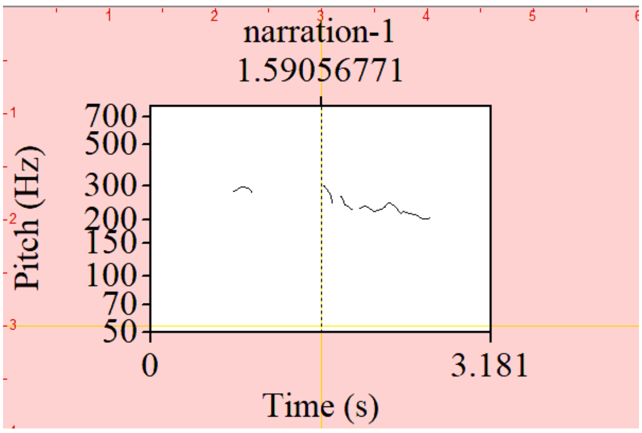

Figure 28: *It's a kangaroo* (1)

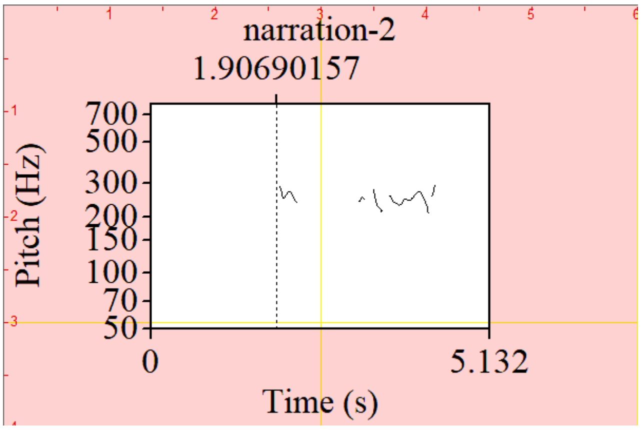

Figure 29: *It's a kangaroo* (2)

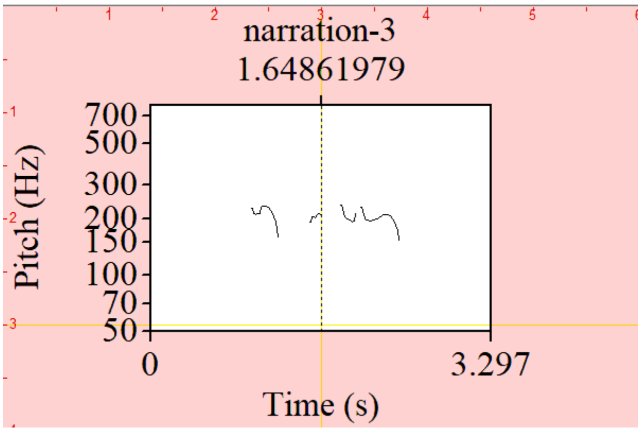

Figure 30: *It's a kangaroo* (3)

The pitch in Figure 28 to 30 is around 300 Hz and above 150 Hz, and the falling pitch at the end of the words is distinctive. The time span ranges from 3 to 5 seconds, indicating that some speakers are speaking slowly. In the word *kangaroo*, the /u/ in the second unstressed syllable, *garoo*, can be either flat or falling. In particular, the vowel of the second unstressed syllable, /u/ has a quite up and down structure as a long acoustic feature.

Next, a sentence, *I'm Jane*, has a distinctive feature on a consonant, /m/, at the position, students tend to take the down pitch; however, a connection at the next word's beginning consonant, /j/ is turned out cutting the tendency. Therefore, the CC consonants, /mj/ in Figure 31 have the same character as Figure 33. However, Figure 32 shows the pitch from /mj/ to /ne/ has a parallel line to the end of sentence.

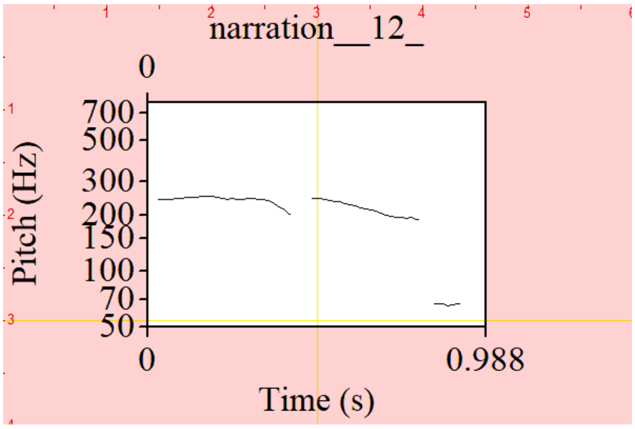

Figure 31: *I'm Jane* (1)

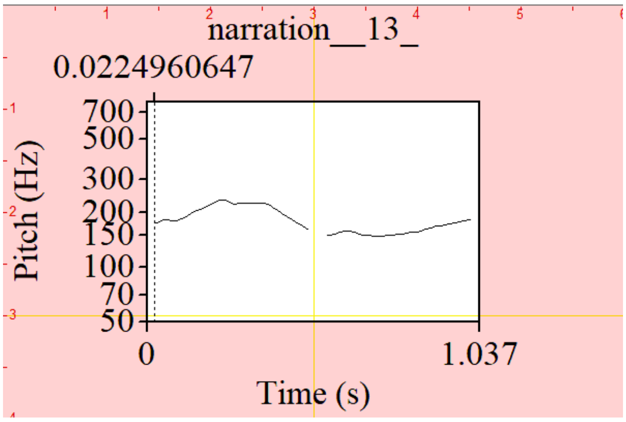

Figure 32: *I'm Jane* (2)

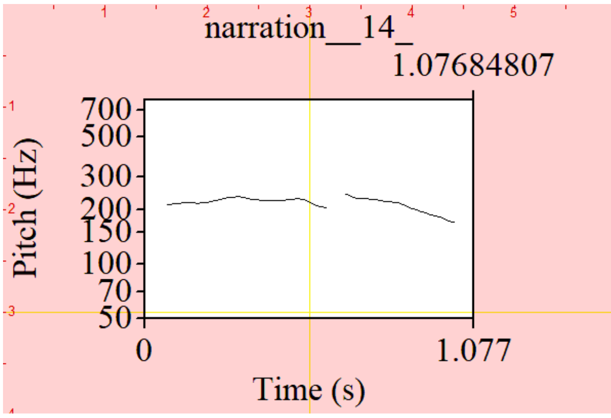

Figure 33: *I'm Jane* (3)

These pitch tones are appeared as the results from recording words and sentences as a vivid structure, where speakers often make a mistake on recognizing the correct intonation and pitch for communication well. It also notes a possibility of linking of consonants and vowels beyond each word's boundary with a plan.

4. Discussion

How can researchers recognize students' voice quality have each tendency? Using a scientific tool of Praat, we get the digitized sound that brings us a tendency from an articulation from many viewpoints like voice onset time, waveforms, spectrum, formants and pitch. Then, it illustrates that syllable distinctions exist in acoustic focuses from consonants and vowels. Therefore, a typical argument is that an experiment on acoustics should be conditional on an inference in choosing consonants and vowels at initial and final positions in syllables from these pilot research for more profitable demonstrations. This has a next experiment with recording words as *sing*, *angry*, and *cleaning* with a nasal of /ŋ/. The initial and final positions on syllables on a relation between advanced consonants' linking in CCC as *street* and vowels should be clarified with the articulation as phonetics. Moreover, words as *laboratory* and *glamorous* with complicated syllables having a morpheme will note that the views of acoustics indicate actual wrong articulations. Then, it needs a task management for building up an experiment on phonetics [16].

Through functions of Praat, the speed on voice onset time has been measured statistically and the result is judged as an existence of distribution from the averages; however, the analysis of spectrum would draw on the same tendency for having a pause between /ŋ/ and /k/ and showing an intonational connection between /ŋ/ and /g/. Therefore, it focuses on the following consonant feature's phonetic conditions of a distinction of plosive voiceless /k/ and voiced /g/ [11]. Next, the results from the word, *kangaroo* note that there is not a pause between the first unstressed syllables, *kan* and a stressed syllable, *garoo* of two syllables. The applied acoustic features on waveforms convey an inference that a syllabic distinction has a differing position by an articulation of vowels from a connection of consonants. Finally, an intonation and pitch are one of quite dominant structures in expressing feeling as communication. This point is deeply related to a mistaken interaction. This experiment also indicates that visual images on pitch suggest that we can see a problem in an intonation from up and down pitch structures on sentence level's sound analyses. The intensity also has a benefit to reveal a speech clearly [17].

A research on foreigners evaluates the relevance of the Accentual Phrase (AP) in French[18]. A corpus of 22 sentences recorded from four native speakers was analyzed for acoustic properties. Results indicated that APs are often realized in the pattern, characterized by peaks with higher F0 values and greater duration. A perceptual investigation showed that listeners could accurately segment sentences into APs. The findings suggest that the perception of an AP as a unit is influenced by tonal, durational, and dynamic features. Moreover, an experiment examined the possibility of word-final F0 rise without vowel lengthening in French [19]. Results indicate variability based on speaker, syllable structure, and syllable number, showing that final rise can occur without lengthening in the phrase-medial final syllable of a noun, unlike the phrase-final syllable, which signals a prosodic boundary through lengthening. The study also explored the duration of infra-syllabic components.

Hopefully, this report suggests that an experiment having recording words and sentences should be organized more carefully.

5. Concluding Remarks

This paper checked that recorded audios reveal a tendency of syllabic construction and covered that visual designs of spectrum and pitch illustrate that speakers make each articulation focus.

Reference

[1]. Zee, Eric: Duration and intensity as correlates of F 0, *Journal of Phonetics* (1978) 6, 213-220.

[2]. Barreda, S. & K. Predeck: Inaccurate but predictable: Vocal-tract length estimation and gender stereotypes in height perception, *Journal of Phonetics* 102 (2024).

[3]. Boersma, D. & P. Weenink: *Praat:* Doing phonetics by computer version 5.3.14. Received from http://www.praat.org

[4]. Zhang, C., K. Jepson, G. Lohfink, A. Arvaniti: Comparing acoustic analyses of speech data collected remotely, *J. Acoust.Soc. Am.* 149, 3910-16, (2021).

[5]. Larkey, L. & W. Strange: Perception of synthetic nasal consonants in initial and final syllable position, *Perception & Psychophysics*, (1978), Vol. 23 (4), 299-312.

[6]. Stone, M.: Evidence for a rhythm pattern in speech production: observations of jaw movement, *Journal of Phonetics* (1981) 9.

[7]. Walsh, T. & F. Parker: Vowel length and 'voicing# in a following consonant, *Journal of Phonetics* (1981) 9, 305-308.

[8]. 0 'Kane, D.: Manner of vowel termination as a perceptual cue to the voicing status of postvocalic stop consonants, *Journal of Phonetics* (1978) 6, 311-318.

[9]. Zee, Erik: Effect of vowel quality on perception of post -vocalic nasal consonants in noise, *Journal of Phonetics* (1981) 9, 35-48.

[10]. Abramsont, A. S. & L. Liskert: Voice-timing perception in Spanish word-initial stops,*Journal of Phonetics*, (1973).

[11]. Henning, R. & A. Jongman: *Phonetics, transcription, production, acoustics, and perception.* Wiley Blackwell (2020).

[12]. O'Connor, J.D. & J.L.M. Trim: Consonant, and syllable-a phonological definition, *Word*, Vol. 9 (2), (1953).

[13]. Caramazza, A., R. Capasso, & G. Miceli: Separable processing of consonants and vowels, *Nature*, Vol. 403 (27), (2000).

[14]. Samuel, G: Speech perception, *Annu. Rev. Psychol, 62*. (2011).

[15]. Ladefoged, P. & S. F. Disner: *Vowels and consonants*. Wiley Blackwell (2012).

[16]. Poncetti, N., E. Parizet, E. Galy& P. Chevret: Effect of hearing impairment on fatigue and annoyance caused by task-irrelevant speech: A working day simulation in an open-plan office, *Applied acoustics, 243*, (2026).

[17]. Kodali, M., A. Luna, S. Kadiri& S. Narayanan: Automatic classification of vocal intensity categories from amplitude-normalized speech signals by comparing acoustic features and classifier models, *Speech*

communication, 174, (2025).

[18]. Rolland, G. & H. Lœvenbruck: Characteristics of the Accentual Phrase in French:an Acoustic, Articulatory and Perceptual Study, *Proceedings of Speech Prosody,* (2002).

[19]. Nakata, S. & Yo. Meynadier: Final accent and lengthening in French, *Conference on Speech Prosody,* (2008), Brazil.

Acknowledgments: This work was supported by JST SPRING, Japan Grant Number JPMJSP2149.