Linguistic Portfolios

Volume 9 Article 10

2020

The Acoustic Phonetic Correlates of Lexical Stress in Japanese-Accented English

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

Linguistic Portfolios–ISSN 2472-5102 –Volume 9, 2020 |120

THE ACOUSTIC PHONETIC CORRELATES OF LEXICAL STRESS IN JAPANESE-ACCENTED ENGLISH

ETTIEN KOFFI

ABSTRACT

According to existing suprasegmental typologies, languages fall into three broad categories: accent, tone, and pitch-accent. Japanese has long been viewed as the poster child of the latter. The fact that Japanese and English (accent language) belong to two different prosodic systems raises three important questions:

1. *Since Japanese is suprasegmentally different from English, can L2 speakers produce English lexical stress intelligibly?*
2. *Which acoustic correlate do they rely on to encode lexical stress in English?*
3. *Does the prosodic strategy used interfere with intelligibility?*

The current study examines these issues by measuring the stress bearing units in <Stella, maybe, brother, also, plastic, Wednesday, station> produced by 10 Japanese L2 speakers of English. Three acoustic correlates of lexical stress (F0, intensity, and duration) are ranked. Just Noticeable Difference (JND) thresholds based on the mean values of these correlates reveal that Japanese speakers of English encode and rank lexical stress as follows: F0 (85.71%) > Intensity (42.85%) = Duration (42.85%).

Keywords and phrases: Lexical Stress of Japanese English, Acoustic Correlates of Lexical Stress, F0/Pitch and Lexical Stress, Intensity and Lexical Stress, Duration and Lexical Stress, Just Noticeable Difference (JND) and Lexical Stress

1.0 Introduction

Hyperboles are commonly used to emphasize the role that lexical stress plays in intelligibility. For example, Field (2005, p. 402) states that "research evidence suggests that suprasegmentals play a more prominent role than segmentals" even though his own findings show that "incorrect misplacement of lexical stress is, relatively speaking, quite small: affecting only around 8% of content words if every word were misstressed," (p. 417). In this paper, an acoustic phonetic methodology based on the Critical Band Theory (CBT) and Just Noticeable Difference (JND) thresholds are used to gauge the intelligibility of Japanese-accented English words. The investigation proceeds in six stages. The first provides a brief overview Japanese suprasegmental system. The second gives some background information about the participants. The third proposes an instrumental definition of lexical stress based on JNDs. Thereafter, F0, intensity, and duration measurements are used to respectively analyze the data. The final sections rank the three correlates, assess intelligibility, and draw pedagogical implications.

2.0 Phonological and Phonetic Parameters of Pitch-Accent Languages

Kawahara (2016, p. 446) has accurately noted that the term "pitch-accent" as used in the phonological literature is semantically ambiguous. Some use it in reference to intonation patterns à la Bolinger (1958). This is NOT the meaning of pitch-accent considered in this paper. Others use the term in relation to lexical stress. This IS the meaning of pitch-accent pursued in this paper. In the latter sense, Japanese pitch-accent is characterized by three phonological and three acoustic phonetic parameters, as displayed in Table 1:

Significance of Parameters	**Pitch-accent**	**Accent (English)**
Pitch is phonemically contrastive	+	–
Pitch is lexically predictable	+	+
Length is phonemically contrastive	+	–
F0/pitch is a robust acoustic cue	+	+
Intensity is a robust acoustic cue	–	+
Duration is a robust acoustic cue	–	+

Table 1: Significant Parameters

These parameters have been well discussed in the literature (Levi 2005, pp. 73-75, Burnham et al. 2014, pp. 1-2, Beckman and Pierrehumbert 1986, Sugiyama 2012) to name only a few. That pitch is phonemically contrastive in Japanese words is illustrated by the following words from Kawahara (2016, p. 447):

1. [ámè] (rain) vs. [àmé] (sweets)
2. [háshì] (chopsticks) vs. [hàshí] (bridge)
3. [sákè] (slamon) vs. [sàké] (alcohol)
4. [kákì] (oyster) vs. [kàkí] (persimmon)
5. [káku] (core) vs. [kàkú] (rank)

That pitch is lexically predictable is also demonstrated by the fact that, according to Kawahara (2016, p. 448), only 14% of Japanese words fit the pattern above. For the remaining 86%, lexical stress can be deduced from stress assignment rules. Sugiyama (2012, p.9) said as much when comparing Japanese and tone languages, “Unlike typical tone languages such as Chinese, in which each syllable is specified for a tone, in Japanese the tone patterns are specified for each word, regardless of the number of syllables.” That length is phonemic in Japanese is also so widely attested and verified that it does require further elaboration, except to note that Kitahara (2001) provides various duration measurements for monomoraic and bimoraic syllables.

As for the acoustic phonetic parameters, there is a very widespread consensus in the literature that pitch-accent languages encode lexical stress by relying overwhelmingly on pitch/F0. According to Sugiyama (2012, p.9) the same goes for Japanese:

> In Japanese, pitch accent manifests itself most consistently and clearly in F0. Other acoustic properties such as intensity and duration are normally not found to be correlated with accent. Duration is not likely to be a correlate of accent because length is phonemic in Japanese. Intensity was once believed to be a correlate of pitch accent, but its status is not certain.

The most noteworthy difference between accent and pitch-accent languages is that, whereas the former can use any of the three acoustic correlates to encode lexical stress, pitch-accent languages rely overwhelmingly on F0. For this reason, it is hypothesized in this paper that Japanese L2 speakers of English would overwhelmingly rely on F0 to encode lexical stress than using intensity or duration. Let's proceed with the data obtained from 10 Japanese participants to see if this prediction is borne out.

3.0 Data Analysis, Participants, and Annotation Procedures

The data for the analysis was obtained from the Speech Accent Archive (http://accent.gmu.edu/howto.php). Ten Japanese speakers of English (five males and five females) read the elicitation paragraph containing the seven disyllabic words in Table 2. According to the transcription in *Oxford Advanced Learner's Dictionary (OALD 2000)*, all the words have a trochaic stress pattern, i.e., the penultimate syllable (in bold) is stressed.

N0	Word	IPA
1.	St**é**lla	[ˈst**ɛ**lə]
2.	m**á**ybe	[ˈm**e**bi]
3.	br**ó**ther	[ˈbr**ʌ**ðər]
4.	**á**lso	[ˈ**ɔ**:lso]
5.	pl**á**stic	[ˈpl**æ**stɪk]
6.	W**é**dnesday	[ˈw**ɛ̃**nzde]
7.	st**á**tion	[ˈst**e**ʃn̩]

Table 2: IPA Transcription of Disyllabic Words

Sociometric data on the participants tell us that their mean age of onset of English is 12.4 years, that is, they started learning English as a foreign language in Japan between the ages of 11 and 15. Their mean age is 28.9. The youngest participant was 20, while the oldest was 49. Their mean length of residency in the USA or in any other the Inner Circle Countries, i.e., countries where English is the native language of the vast majority of the population is 2.6. One participant had never left Japan but the remaining nine have visited or lived abroad for 8.7 years. Collectively, the participants produced 420 tokens (7 words x 2 syllables x 10 participants x 3 parameters). The annotation procedures used to measure their data are illustrated by Figure 1. The top portion displays the waveform of <plastic>. The middle section highlights the pitch track. The bottom part shows the annotations and the measurements. The focus is on the syllable nuclei <a> in <plas> and <i> in <tic> because nuclei are stress bearing units. This same procedure was used to collect all 420 tokens.

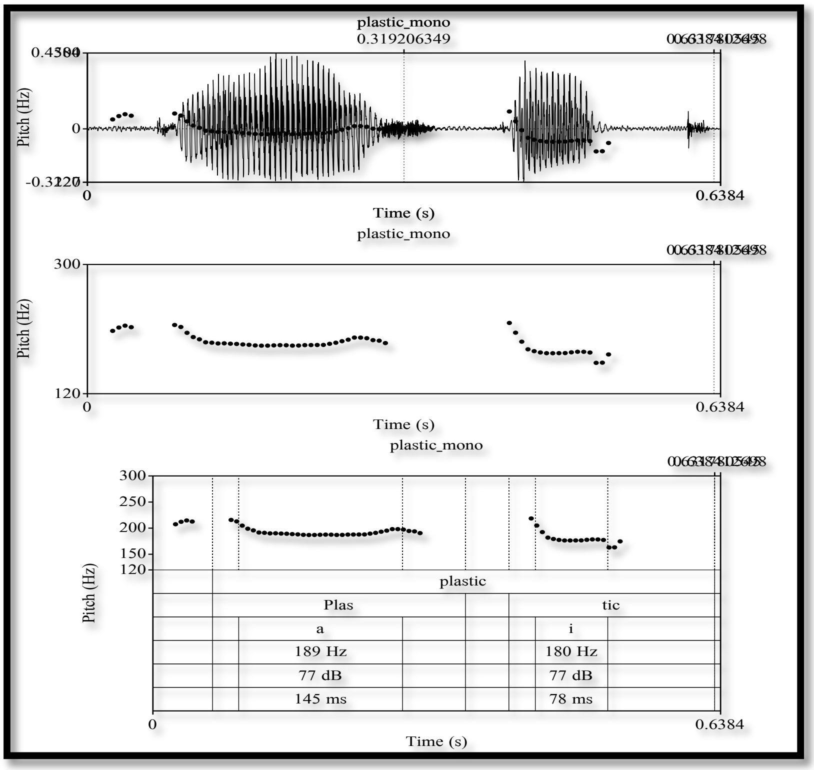

Figure 1: Annotation of "Plastic" by Japanese 3F

4.0 A New Definition of Lexical Stress

The definitions of lexical stress are most commonly stated as follows:

> In many languages, certain syllables in a word are louder, slightly higher in pitch, and somewhat longer in duration than other syllables in the word. They are stressed syllables (Fromkin et al. 2017, p. 205).

This is a good definition; however, it is not without problems. For one, the definition is impressionistic, that is, it relies on the auditory sensations of the hearer. We all know that auditory illusions are real as noted by Katz (2013, p. 179). Baken and Orlikoff (2000, p. 1-2) also write that "The ear is too easily fooled." Secondly, impressions are subjective. This explains why people do not necessarily agree on terms such as "louder," "higher in pitch," and "longer in duration." We, therefore, need a definition of lexical stress that is objective, quantifiable, and operationalizable. Koffi (2019, p. 160) has provided such a definition based on CBT and JND thresholds. Here it is:

> A strong/stressed syllable is one whose F0 is ≥1 Hz higher, whose intensity is ≥3 dB louder, or whose duration is ≥10 ms longer than any other syllable(s) within the same word.

This definition is a summation of more than half a century of psychoacoustic research, experimentations, and findings. Koffi (2017) and Koffi (2018) has explained the rationale for appealing to CBT for a definition of lexical stress and the auditory perception of suprasegmentals and autosegmentals. There is no reason repeat the same arguments here, except to mention that one of the leading proponents of CBT was awarded a Nobel Prize for his discoveries of how the basilar membrane transduces speech signals for the auditory perception of sounds. The JNDs mentioned earlier in the definition of lexical stress are further clarified in the sections below and used to rank the acoustic correlates lexical stress in Japanese-accented English.

4.1 Significance of the F0 JND in Encoding Lexical Stress

Numerous psychoacoustic experiments going back to more than 60 years have confirmed that the human ear is very sensitive to F0 fluctuations. Young (201, p. 609) summarizes the evidence as follows:

> The perception of frequency is called pitch. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different pitch from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more.

Stevens (2000, p. 228) adds that the ≥1 Hz JND works for all normal speech events with intensity levels between 60 to 80 dB. Phoneticians have applied this JND to pitch analysis as far back the 1970s (Lehiste 1970, p. 64 and Gandour 1978, p.57) among others. When this JND is applied to the data in Table 3, it yields the following results.

Words	St**é**lla		m**á**ybe		br**ó**ther		**á**lso		pl**á**stic		W**é**dnesday		st**á**tion	
F0	ste	la	may	be	bro	ther	al	so	plas	tic	wenz	day	sta	stion
Japanese 2F	238	278	235	270	220	240	237	313	230	256	212	251	208	82
Japanese 3F	198	116	221	198	194	186	173	173	189	180	220	245	176	116
Japanese 4M	130	162	146	141	120	111	127	127	121	118	118	119	114	101
Japanese 5F	210	178	232	241	214	206	267	271	225	228	221	186	205	179
Japanese 8M	113	94	143	118	98	91	198	202	114	110	119	90	84	81
Japanese 9M	139	116	151	142	134	133	182	191	170	158	147	147	117	116
Japanese 10F	173	162	164	161	158	152	197	230	178	174	166	151	82	78
Japanese 11F	190	170	212	221	183	171	226	239	202	196	189	185	82	74
Japanese 12M	99	74	113	116	105	107	96	178	104	101	96	111	95	74
Japanese 13M	104	122	165	107	127	92	188	139	146	139	155	117	130	105
Mean	**159**	**147**	**178**	**171**	**155**	**148**	**189**	**206**	**167**	**166**	**164**	**160**	**129**	**100**
St. Dev.	48	57.5	43.1	57.4	44.9	51	50.2	57.9	45.3	51.1	45.2	55.6	49.6	32.1

Table 3: F0 Measurements

The Japanese speakers relied overwhelmingly on F0 to encode lexical stress in 6 out of the 7 words (85.71%). This verifies the hypothesis formulated earlier, namely, that since Japanese is a pitch-accent language, non-native speakers would transfer the lexical stress strategy of their L1 into their L2. This finding is also consistent with Kondo (2009).

4.2 Significance of Intensity in Encoding Lexical Stress

The JND for a differential perception of intensity is universally accepted to be ≥ 3 dB (Stevens 2000, p. 225). It is found in the sensitivity specifications of many audio products and many sound level meter apps. This JND is endorsed by NIOSH (National Institute for Occupational Safety and Health) and many reputable national and international organizations.[1] When this JND is applied to the data, it leads to the following results:

Words	St**é**lla		m**á**ybe		br**ó**ther		**á**lso		pl**á**stic		W**é**dnesday		st**á**tion	
Intensity	ste	la	may	be	bro	ther	al	so	plas	tic	wenz	day	sta	tion
Japanese 2F	78	82	75	76	73	74	77	74	73	71	73	73	70	64
Japanese 3F	84	74	80	76	77	77	77	78	77	77	80	81	75	69
Japanese 4M	81	71	80	80	80	79	81	80	78	76	80	78	78	68
Japanese 5F	70	67	76	70	80	71	77	74	78	70	74	72	68	58
Japanese 8M	80	72	76	75	75	72	79	77	75	69	75	74	69	61
Japanese 9M	80	71	82	79	75	74	84	83	80	77	79	79	69	63
Japanese 10F	81	73	77	75	76	77	84	83	79	76	75	76	76	68
Japanese 11F	72	67	73	69	72	70	74	72	70	71	72	73	69	60
Japanese 12M	69	54	71	70	73	72	76	66	68	64	70	63	67	58
Japanese 13M	74	67	82	80	78	72	78	77	79	76	82	74	78	70
Mean	**76**	**69**	**77**	**75**	**75**	**73**	**78**	**76**	**75**	**72**	**76**	**74**	**71**	**63**
St. Dev.	5.3	7.1	3.7	4.1	2.8	2.9	3.3	5.1	4.1	4.3	4	4.9	4.3	4.6

Table 4: Intensity Measurements

The participants relied minimally on intensity to encode lexical stress. They did so in three out of seven words (42.85%) in <Stella>, <plastic>, and <station>. In <maybe>, <brother>, <also>, and <Wednesday> intensity was not a factor. This finding agrees with Sugiyama (2012, p.16) in that Japanese speakers hardly rely on intensity to encode lexical stress in their own language.

4.3 Significance of Duration in Encoding Lexical Stress

Gunnar (1960, p. 233) credits Fletcher with having discovered the JND for duration in 1929. A sound is perceived as being longer than another if the durational difference between them is ≥10 ms. Moore (2007, p. 468) summarizes these duration experiments as follows, "Duration discrimination has typically been studied by presenting two successive sounds which have the same power spectrum but differ in duration. The subject is required to indicate which sound had the longer duration." Experiments too many to mention here have confirmed the reliability of this threshold, including Hirsh (1959) and Stevens (2000, p. 228). When this JND is used to examine the data in Table 5, it yields the following results.

[1] OSHA (Occupational Safety and Health Administration) and the audiology profession use the JND ≥ 5 dB to screen for auditory acuity in healthcare and occupational safety. The ≥ 3 dB JND is for every day auditory perception of intensity levels. The two JNDs serve two different purposes.

Words	St**é**lla		m**á**ybe		br**ó**ther		**á**lso		pl**á**stic		W**é**dnesday		st**á**tion	
Duration	ste	la	may	be	bro	ther	al	so	plas	tic	wenz	day	sta	tion
Japanese 2F	82	158	181	95	85	99	61	125	97	81	67	111	103	61
Japanese 3F	69	63	139	98	110	73	55	98	145	78	92	158	122	77
Japanese 4M	69	86	139	98	82	97	54	105	81	83	72	142	97	96
Japanese 5F	83	74	162	86	111	86	72	76	87	58	80	237	153	87
Japanese 8M	74	98	69	40	61	73	75	52	90	31	47	40	82	42
Japanese 9M	68	62	62	47	60	27	61	57	37	37	56	57	88	41
Japanese 10F	108	92	144	189	77	57	50	47	52	31	86	163	77	111
Japanese 11F	50	68	179	144	72	77	127	164	55	57	84	145	90	45
Japanese 12M	127	73	114	109	72	82	89	51	61	57	91	62	114	59
Japanese 13M	42	67	100	107	93	85	44	37	51	61	103	188	95	52
Mean	**77**	**84**	**128**	**101**	**82**	**75**	**68**	**81**	**75**	**57**	**77**	**130**	**102**	**67**
St. Dev.	25.1	28.7	41.9	42.9	17.9	20.9	24.3	40	31.5	19	17.3	62.7	22	24.4

Table 5: Duration Measurements

The participants relied on duration to encode lexical stress in only 3 out 7 words (42.85%). They do so in <maybe>, <plastic>, and <station>. In the remaining words, duration does play any role. Again, Sugiyama (2012, p.9) reports that since duration is phonemic in Japanese, native speakers hardly rely on it to encode lexical stress. This also may explain why it is used marginally to encode lexical stress in Japanese-accented English.

4.4 Results and Correlate Ranking

It stems from the preceding analyses that Japanese L2 speakers of English rank (subconsciously, of course) the acoustic correlates of lexical stress as follows:

F0 (85.71%) > Intensity (42.85%) = Duration (42.85%)

This shows that they overwhelmingly rely on F0 to encode lexical stress. Intensity and duration are used only sparingly. The very fact that they are used minimally can be construed also as evidence of prosodic transfer from the L1. Sugiyama (2012, p.16, 18) notes that even in Japanese one cannot say categorically that intensity and duration are not used at all:

> Some studies show some possible correlation between accent and duration or intensity, even though it is not strong. … Based on the data from Beckman (1986) and Weitzman (1970), it may be too strong a claim that there are no correlates other than F0 for pitch accent. Intensity and duration are correlated with accent to some extent, but the correlations are not strong.

Levi (2005, p. 92) makes a similar observation about the role of intensity and duration in Japanese. Even so, the ranking uncovered in this paper aligns perfectly with previous findings of Japanese pitch-accent and Japanese-accented English. Kondo (2009, p. 105), for instance, shows that Japanese speakers rely mostly on F0 to encode lexical stress in English, but that they do not make much use of intensity and duration. She sees this as an "influence of Japanese phonology."

5.0 Intelligibility Assessment

We are now in a position to answer our last research question, i.e., does the prosodic strategy used by Japanese speakers to encode lexical stress interfere with intelligibility? The answer to this question is a resounding and an emphatic no. Previous studies, including Fry's (1958, p.151) seminal paper ranked the acoustic correlates of lexical stress in English as F0 > Duration > Intensity. Many subsequent studies have also ranked F0 first. My study of the 10 native speakers of American English who produced the same exact words in Table 2 as the 10 Japanese L2 speakers shows that they encode lexical stress as follows: Intensity (85.71%) >F0 (57.14%) > Duration (28.57%). A similar ranking in which intensity is the top strategy for encoding lexical stress in English is reported in Kochanski et al. (2005). Regardless of which correlate comes on top, each of the three acoustic correlates can be used equally to encode lexical stress. Consequently, so long as the segments onto which suprasegments are anchored are produced intelligibly, Japanese-accented English cannot be misunderstood if the talkers rely on overwhelmingly on F0 to encode lexical stress.

6.0 Summary and Pedagogical Implications

Kondo (2009) contains a number of statements which seem to imply that there is something amiss with relying only on F0 to encode lexical stress. For example, she writes on page 107 that the eight Japanese who participated in the study were "fluent in English but none them were bilingual." How can this be? In a 2018 study involving 72 Japanese L2 speakers of English who were divided into beginners and advanced learners, Kondo and two other co-authors contend that advanced Japanese learners relied on duration more than F0 or duration to encode lexical stress (Figure 4, p. 12) whereas beginners relied more on F0 than duration and intensity (Figure 3, p. 12). The implication seems to be that the more proficient a Japanese speaker becomes, the more likely he/she is to use duration cues to encode lexical stress. Coincidentally, the Advanced Japanese speakers used duration 39.4% of the time to encode lexical stress like the native English speakers in their study (39.1%). The insinuation that when Japanese speakers become "advanced," they rely on duration to encode lexical stress is not supported by the data presented in this study nor by a large amount of measurements that I have collected over the years from highly proficient undergraduate and graduate Japanese exchange students at my university. When they come to my university they enroll directly in courses with domestic students because their TOEFL scores are very high and also because my university has an agreement with Akita International University where English is the medium of instruction. These exchange students are fully bilingual. Yet, they still encode lexical stress in English by relying overwhelmingly on F0.

ABOUT THE AUTHOR

Ettien Koffi, Ph.D. linguistics, teaches at Saint Cloud State University, MN. He is the author of four books and author/co-author of several dozen articles on acoustic phonetics, phonology, language planning and policy, emergent orthographies, syntax, and translation. His acoustic phonetic research is synergetic, encompassing L2 acoustic phonetics of English (Speech Intelligibility from the perspectives of the Critical Band Theory), sociophonetics of Central Minnesota English, general acoustic phonetics of Anyi (a West African language), acoustic phonetic feature extraction for application in Automatic

Speech Recognition (ASR) and Text-to-Speech (TTS), and voice biometrics for speaker verification. He can be reached at enkoffi@stcloudstate.edu.

References

Baken, R. J. and R.F Orlikoff. (2000). *Clinical Measurement of Speech and Voice*. Second Edition. San Diego, CA: Singular Thomson Learning.

Bolinger, D. L. (1958). A Theory of Pitch Accent in English. *Word* 14 (2-3):109-149.

Beckman, M. E. and Pierrehumbert J. B. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook 3*: 255-309.

Field, J. (2005). Intelligibility and the Listener: The Role of Lexical Stress. *TESOL Quarterly* 39 (3):399-423.

Fromkin, V., R. Rodman., and N Hyams. (2017). *An Introduction to Language*. 11th edition. Boston, MA: Cengage Learning.

Fry, D. B. (1958). Experiments in the Perception of Stress. *Language and Speech* 1: 126-152.

Gandour, J. (1978). The Perception of Tone. *Tone: A Linguistic Survey*, pp. 41-76, ed. by Victoria Fromkin. New York: Academic Press

Gunnar, F. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton & Co.-S-Gravenhage.

Hirsh, I. J. (1958). Auditory Perception of Temporal Order. *Journal of the Acoustical Society of America* 31 (6):759-767.

Katz, W. F. (2013). *Phonetics for Dummies*. Hoboken, NJ: John Wiley & Sons, Inc.

Kawahara, S. (2016). The Phonology of Japanese Accent. *De Gruyter/TCS*, pp. 445-492.

Kitara, M. (2001). *Category Structure and Function of Pitch Accent in Tokyo Japanese*. Indiana University, Bloomington: Ph.D. Dissertation.

Kochanski, G., E. Grabe, J. Coleman, and B. Rosner. (2005). Loudness Predicts Prominence: Fundamental Frequency Lends Little. *Journal of the Acoustical Society of America* 118 (2):1038-1054.

Kohishi, T., J. Yun, and M. Kondo. (2018). Acoustic Correlates of L2 English Stress – Comparison Japanese English and Korean English. *Phonetics and Speech Science* 10 (1):9-14.

Kondo, M. (2009). Is the Acquisition of L2 Phonemes Difficult? Production of English Stress by Japanese Speakers. *Proceedings of the 10th Generative Approaches to Second Language*, pp. 105-112, ed. by Melissa Bowles et al.. Somerville, MA: Cascadilla Proceedings Project.

Koffi, E. (2019). A Template Model Account of Lexical Stress in Arabic-Accented English. In J. Levis et al. (Eds). *Proceeding of the 10th Pronunciation in Second Language Learning and Teaching Conference*., pp. 158-167. Ames, IA: Iowa State University.

Koffi, E. (2018). Differential Analysis of Lexical Pitch in Accent and Tone Languages. *Linguistic Portfolios* 7: 110-131.

Koffi, E. (2017). The New Paradigm in Tone Analysis: The Contribution of the Critical Band Theory. *Linguistic Portfolios* 6: 147-165.

Ladefoged, P. (1996). *Elements of Phonetics*. Chicago: The University of Chicago Press.

Lehiste, I. (1976). Suprasegmental Features in Speech, pp. 225-239. In *Contemporary Issues in Experimental Phonetics*, ed. by Norman Lass. New York: Academic Press.

Levi, S. (2005). Acoustic Correlates of Lexical Accent in Turkish. *Journal of the International Phonetic Association* 35 (1): 73-97.

Moore, B. (2007). Psychoacoustics, pp. 459-501. In *Springer Handbook of Acoustics*, ed. by T. D. Rossing. New York: Springer Science+Business, LLC

Stevens, K. (2000). *Acoustic Phonetics*. Cambridge, Massachusetts: The MIT Press.

Sugiyama, Y. 2012. *The Production and Perception of Japanese Pitch Accent*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Young, H. (2011). *College Physics*. 9th edition. Addison-Wesly: NY.