

External and Internal Articulatory Gestures Made by Japanese Speakers of English

TAMAKI, Takatoshi^{*}, MENISH, Marc^{**}, DUFF, Brian^{***}

Abstract This research was based on the assumption that certain consistencies in external and internal articulatory gestures by upper-intermediate and advanced learners of English might become evident under suitable experimental conditions. The authors of this study were able to identify several key markers in both categories of gesticulation and are hopeful that these findings may assist further exploration in this and similar linguistic applications.

Keywords Articulatory gesture, head movement, facial expression, jaw movement, intonation, pronunciation, tongue position, morae, pause, intonational group, rhythmic pattern, stress patterns

Origins of Study

In April of 2016, the authors began constructing a laboratory of sorts in the corner of a research office. In addition to a black backdrop, there were several important pieces of equipment provided by the Office of Information Science at Aoyama Gakuin University that enabled our research efforts. A high definition camera with advanced slow motion capabilities, a high aperture lens for indoor recording, a television prompter for frontal reading by the subjects, and various computer programs allowed us to capture, edit and analyze key data sets.

Over 100 subjects- including students, teachers and lay people have been videotaped for analysis. Within this group, a sub-set of approximately 45 college students of Japanese nationality become the primary group for close analysis. Focusing on one particular group with a similar language traits allowed us to compare results with higher accuracy.

Part 1: Upper-intermediate and Advanced Learners of English

This first experiment is aimed at observing and analysing how our Japanese subjects demonstrate their memorised speech in English. It is also to examine whether there are features of articulatory gesture characteristic of them. We will look at both external and internal articulatory gestures, the former ones in particular, because the Japanese in general are said to be notoriously poor in them.

Student Videos 1 is analysed on the basis of the following criteria: A) head movement; B) facial expressions; C) lip (and jaw) movement; D) pronunciation, and E) rhythm and intonation. The former three criteria may indicate external articulatory gestures and the latter ones internal articulatory gestures postulated under perceptual observation.

1. External articulatory gestures

1.1 Head movement:

The movement of the subjects' heads based on how they correspond to word stress:

The head movement related to word stress is an act of nodding and its degree is generally small. It is made by roughly 75% [32] of all the subjects [43] at the stressed syllable on every word. 18 subjects among them make a big movements with their heads at the last word in a phrase unit (see 2.2.1 (b)).

1.2 Facial expression:

The expression on the subjects' faces when looking directly at the camera.

The facial expression of nearly all the subjects looks serious partly because they are not used to conducting themselves

* Professor of Linguistics (ret.), Aoyama Gakuin University

** Professor of Media, Aoyama Gakuin University

*** Professor of Business, Aoyama Gakuin University

in front of the camera and slightly embarrassed as a result.

1.3 Lip (and jaw) movement:

The way the lips show the characteristics of a particular sound pronounced by the subject. The movements of the lower lip and jaw should be observed together as both the organs of speech generally cannot move separately.

Except the closure of the upper and lower lips for bilabial sounds, the lip movement is generally not conspicuous. Referring to the /f/ and /v/, they are thought to be pronounced with strong lip-rounding in English. Visible lip-rounding for them is not observed among the subjects.

2. Internal articulatory gestures

2.1 Pronunciation:

the contrast between the voiceless alveolar fricative /s/ and the voiceless palate-alveolar fricative /ʃ/:

/ʃ/ is substituted for /s/ when /s/ and /t/ are put together to become a syllable as in 'decision' (3, 4, 9, 14, 16, 17, 19, 20, 21, 22, 24, 27, 28, 33, 36, 37, 39, 42);

the contrast between the voiceless labiodental fricative /f/ and the voiceless bilabial fricative Jpn /ɸ/, that between the voiced labiodental fricative /v/ and the voiced bilabial plosive /b/:

/f/ is replaced with Jpn/ɸ/, and /v/ is replaced with /b/ (3, 4, 5, 9, 11, 12, 16, 17, 18, 20, 21, 22, 23, 24, 34, 37, 39, 40, 43);

the labiovelar approximant /w/ v the Jpn vowel /u/:

no subjects satisfactorily pronounce the /w/ in terms of lip-rounding, the tongue position of the Jpn vowel /u/ is the same as the /w/ but without lip-rounding;

the voiced and voiceless dental fricatives /ð/ and /θ/:

these sounds tend to be pronounced as /dz/ and /s/ (3, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 25, 30, 31, 32, 33, 35, 37, 38, 39, 40, 42);

the vowel /ʌ/, as in 'company' :

pronounced as /æ/ by 27;

'flawed' /ɔ:/:

pronounced as /aʊ/ (9, 10, 30, 31, 36, 40).

Stress pattern:

'optimism' ('-mism' is stressed by 1, 4, 17, 29, 35, 39),

'bankruptcy' ('-rupt' is stressed by 3, 24, 25, 40),

'ugly' ('-ly' is stressed by 33, 35, 37).

2.2 Rhythm and intonation:

2.2.1 Rhythm

(a) Rhythmic pattern

'correct them' / 'start it with' just before the phrase-unit boundary:

the former rhythmic pattern is correctly articulated by 7, 8, 10, 12, 13, 26, 28, 30, 31, 34, 36, 38, 43 whereas the latter one by 1, 2, 27, 29, 32, 33, 39;

'founder conflict' [compound noun (Noun + Noun)], cf the noun phrase (Adjective + Noun):

no subjects follow the expected rhythmic pattern of ● ● ● ● (Also see Part 2, section 3.)

Except that the pronoun in English can be stressed depending on context, it will be unstressed. In terms of rhythm, it generally has both weak and strong forms. Although it is unstressed, it will be realised as a strong form when it is the last word in a phrase before a pause.

The ‘on’ of ‘made early on’ in this case is an adverb, and should be stressed. Nearly all the subjects put stress on every word in each sentence, and therefore it may be most unlikely that they intend to have it stressed only because the ‘on’ is an adverb.

The compound noun consisting of two (or more) words will have the former one stressed. Strictly speaking, it will be stronger in stress than the latter because the element of intonation is added to it. By contrast, in the noun phrase, ‘social conflict’, for example, both the adjective and the noun will be stressed, but the latter will be stronger in stress. The element of intonation is added to it in this case, too.

Some of the subjects (11, 16, 23, 37) articulate their words with syllable-timed (*or* what they call ‘mora-timed’) rhythm, where syllables (*or* morae) tend to occur regularly in timing.

(b) Phrase unit

The phrase unit may be a unit determined by pauses, which is either a word, a phrase, a clause or a sentence as a whole. In running speech it may also be expressed as a tonal group (*or* an intonational group, a phonological phrase) (also see Appendix as a reference). However, phrase units are also able to combine together without pauses. Whether or not there is a pause between phrase units may be decided normally both by the speaker’s interpretation of a context and by their rate of speech. In spontaneous speech phrase units, this was not observed.

2.2.2 Intonation

The majority of subjects adopted a rising intonation at the last stressed word in each phrase unit of a sentence, and a falling one at the last stressed word. Subjects 20 and 43, however, use a rising intonation at the end of the statements.

Each of the following passages shows the overall placement of pauses made by the subjects. They are indicated as /. Each chunk between pauses corresponds in most cases to a phrase unit;

Companies / are like countries in this way. / Bad decisions || made early on / – if you choose the wrong partners / or hire the wrong people, || for example -/- are very hard to correct / after they are made. / It may take a crisis || on the order of bankruptcy / before anybody || will even try to correct them. / As a founder, / your first job / is / to get the first thing right, / because || you cannot || build a great company / on a flawed foundation. /

When you start something, / the first and most crucial decision || you make / is / whom || to start it with. / Choosing a co-founder / is like getting married, / and founder conflict / is just as ugly as divorce. / Optimism / abounds / at the start of every relationship. / It’s unromantic to think soberly / about || what could go wrong, / so people don’t. / But if || the founders develop / irreconcilable differences, / the company / becomes the victim. /

* || indicates the unusual placement of pause.

Considering the head movement, the facial expression in front of the camera, the pronunciation and the articulation rate, the following subjects have for the most part fulfilled the criteria;

7. RM

8. GT

36. RT.

The following subjects are generally good, but not satisfying some of the criteria above;

26. YY (the head movement and articulation rate)

27. DY (the pronunciation)

38. NT (the pronunciation).

3. Discussion

The facial expression demonstrated by the subjects is generally blank or monotonous. Many subjects turned their eyes towards the camera but appear serious without any discernible expression.

In terms of the stress pattern, almost every word is correctly pronounced. Also see above.

English is thought to be a stress-timed language, which means that sentence stresses in a phrase unit tend to occur roughly at the same amount of time. In other words, rhythm in English can be psychologically, not physically, real and empathetically (cf ‘phonetic empathy’ coined by Abercrombie (1967)) felt by the listener. This could also be applied to rhythms in any other languages, and there is no wonder why rhythm in a language is quite often said to play an extremely important role in speech.

Individual words in English contain their own primary stress whether or not they consist of a single syllable or two (or more than two) syllables. Grammatically, words in English can be divided into two categories: content and function words. The former tend to be stressed in a phrasal units. Stresses— that is, sentence stresses— on them play a pivotal role and tend to occur psychologically at regular intervals of time. Among them, the stress on a particular word may also be responsible for being the nucleus of the unit because the word may contain new information or be emphasized as a stylistic feature within it. When speech encounters the nucleus, newsreaders can be seen making such physical gestures as head movements, mainly together with their mouth and/or eyes widening. Some body and/or hand movements may also become involved.

Function words generally are unstressed in a phrase unit. The results of our experiment reveal that each head movement coincides with almost every word, whether a content word or a function word, in a phrase unit, or it does so with the last word in it. The latter case looks quite similar to the following tendency. Many Japanese, newsreaders in particular, tend to nod when they come to the first and/or the last noun or verb in a phrase unit. (The phrase unit in Japanese is determined by the glottal stop, or the glottal stop and a pause (Tamaki, 1998)).

4. Concluding remarks

Close observations suggest that there is among the subjects a marked tendency that every word is pronounced with a (sentence) stress on it. This could result in the rhythmic tactics that may be different from those employed by native speakers of English. They should be substantiated as external articulatory gestures quite unique to the subjects.

In Japanese, the noun phrase consists of a noun followed by a particle and the verb phrase a verb followed by a particle. Nouns can be followed by such particles as ‘wa’, ‘ga’, ‘de’ whereas the verb by such as ‘desu’, ‘da’. The head movement, as mentioned above, often appear at the start of the noun or verb. Instead of this, it is quite often made at the beginning of such a particle as ‘desu’. As we can see, head movements in Japanese speech appear to be inconsistent when considering those in English.

Part 2: Articulatory Gestures by Native Speakers of English

1. Experimental purpose

The main purpose of this second experiment is to analyze external articulatory motions in speech made by native speakers of English within the same age group as that of our Japanese intermediate and advanced learners of English.

Our subjects are two American students, one male and the other female. They were instructed to read aloud the same two passages taken from *Zero to One* (Thiel, et al 2014), adopted in the first experiment. The only difference between the past and present experiments is whether or not the subjects have been told to memorize them beforehand. We examined them on video, and have found out that they just follow each line of the passage on the screen in front of them to read it aloud. We have ignored the eye movement of our subjects because it could not be thought to contribute to external articulatory movement in speech.

2. Data analysis

Based on the first experiment, we are particularly interested in how rhythm in English speech is realized in terms of external articulatory gestures. The scope of the present situation is confined to delivering the content of the passages, in which a wide variety of gestures could not be expected. The common characteristic of the two subjects is the jaw motion demonstrated through the read speech. Both the male and female subjects do not show any sign of other external movements at all except the latter shows her head movements (that is, nods) in part of the latter part of the second passage:

‘As a founder, / your first job / is to get the first thing right, / because you cannot / build
N N N N N N N N N N
a great company / on a flawed foundation.’
N N

(/ indicates pause; N under words shows that nods coincide with the stressed syllables of them; nods cannot be detected in ‘flawed’ and ‘foundation’.)

The jaw motion in speech normally illustrates the vertical movement of the lower jaw although its degree emotionally and/or physically depends on speakers themselves. We have found out that our both subjects lower their jaw slightly more than average when it comes to a content word with its sentence stress. The movements indicate clearly and visibly the unique patterns of rhythm in English.

3. Jaw movement, rhythm and intonation

Throughout the observation of external articulatory gestures, jaw movement is our major source of rhythmic beats in speech. Concerning the female subject, this is not visibly identified at the beginning of the passage she has been instructed to read aloud. We, therefore, must rely on her pitch variations in order to detect rhythmic beats in speech. The beats coincide with syllables that will receive what we call a ‘sentence stress’ in speech. Perceptual prominence or stress in general can be obtained when a syllable is higher in pitch (roughly corresponding to acoustic fundamental frequency), longer in length (acoustic duration), and/or louder in loudness (acoustic intensity). Loudness is said to be quite similar to sonority, which is a natural characteristic of a particular speech sound. Vowels are described as being more sonorous than consonants are. Quite often, a scale of degree of sonority is said not to be regarded as exact measurements (Ladefoged 1993, 245-247). However, a rigorous experiment measuring intensity of all the phonemes realized in speech in three languages, English, Spanish and Quechua proves that sonority corresponds to acoustic intensity (Parker 2008).

Our two subjects demonstrate that rhythmic beats (that is, sentence stress) in speech coincide with content words, which include ‘don’t’ in our passages. However, we have to bear in mind that all the content words in a text do not necessarily possess their rhythmic beats, depending on its context as well as a speaker’s interpretation of it. This is also applied to function words, which are, as a rule, unstressed in speech. Our male subject demonstrates this in his speech. In ‘When you start something, the first and most crucial decision you make is whom to start it with.’, he emphasizes ‘you’ of the underlined clause and places a rhythmic beat on it (Figure 4.2; cf Figure 4.1. See also Section 4.).

Incidentally, one peculiar interpretation of part of the context is demonstrated by the female subject. In the ‘companies are like countries in this way’ passage, ‘...after they are made’ sounds to be cut out from its sentence and added to the next as if it were the beginning clause. The word ‘correct’ ends with a falling intonation to mark the finality. Then, after a pause, the ‘...after they are made’ begins.

Referring to ‘companies’ [●●●] in one passage and ‘optimism’ [●●●] in the other, they are traditionally syllabified into three syllables. The ‘com-’ in ‘com·pa·ny’ and ‘op-’ in ‘op·ti·mism’ are lower in pitch than their latter parts, ‘-pany’ and

‘timism’ respectively. Perceptually, the female subject shows the stress pattern of ‘companies’ like a [●●●] whereas the male subject a [●●].

The story of ‘optimism’ in both subjects is slightly different from that of ‘companies’. Their stress pattern commonly seems to be quite like a [●●]. The female subject is also quite likely to insert an extra vowel [ə] between the [mɪz] and [m] in the ‘-mism’. This can be identified in her spectrogram, and the last peak in the intensity figure also suggest this phenomenon.

Although their tactics towards the both words are uniquely common to them, the ‘com-’ and ‘op-’ are accompanied by a slightly lower jaw movement, which may prove that the ‘com-’ and ‘op-’ are primarily stressed.

4. Compound nouns in English

Generally speaking, compound nouns in English are assigned their first noun to be more stressed than the second one. If this rule is also applied to the ‘founder conflict’, it should be schematically depicted as a [●●●].

There are many words in English having their own noun and adjective as a part of speech. As a general rule, when they stand as the first element of a compound noun, they will receive a stronger stress than the latter ones. If they do as an adjective, that is, the first element of a noun phrase, the latter nouns will be stronger in stress than the former ones.

However, the stress pattern of compound nouns is quite tricky because reality in fact does not necessarily reflect our expectations. The followings are some examples: ‘orange juice’ [●●]; ‘lemon squash’ [●●] (cf ‘lemon squeezer’ [●●●]); ‘English teacher’, which will be either [●●●] (a teacher who teaches the English language) or [●●●] (a person coming from England, UK). In the latter of which the ‘English’ is an adjective and should be a noun phrase. A rhythmic beat should be assigned to the [●].

As for ‘founder conflict’, the ‘founder’ itself is either a noun or a verb but it naturally is a noun in this case. Phonotactically, the stress pattern could be a [●●●]. [●●●] is, however, assigned to ‘founder member’ (someone contributing to inaugurating a new organization such as a club, business or school and is one of its first members). One of the co-authors, Blake Masters (Thiel 2014), also adopts this pattern for the ‘founder conflict’ in his recordings. It might be accepted with a rhythmic beat being applied to the ‘found-’ and ‘mem-’ each.

Judging by the external articulatory gestures, the male subject demonstrates that the stressed syllables, ‘found-’ and ‘con-’, of both words are with slight lower jaw motion. When the jaw movement of the former syllable is compared with that of the latter, the former one can be found out to be bigger than the latter. The downward jaw motion of the female subject, however, tells us that she demonstrates the same degree of departure from the neutral position of the jaw in the stressed syllables, ‘found-’ and ‘con-’, of both words. The ‘neutral’ means that the jaw as well as speech organs muscles are ready for articulation, not relaxed.

5. Acoustic aspects of speech

Perceptual observations on the articulation of our two subjects are now compared with acoustic examinations of their speech. Which will lead us to know the differences and similarities between them. The followings are some aspects of acoustic phonetics relevant to this matter.

For acoustic analysis of the two subjects’ read speech the soundtrack of the original video on PC was transferred into .wav format using Audio Creator LE (Cakewalk 2008). Then the waveform was edited to single out appropriate chunks into separate .wav format files with DigiOnSound X (DigiOn 2016). Those chunks of the waveform were further analyzed with the acoustic analysis software Speech Analyzer (SL International 2012).

Each of the figures (Figures 1.1, 1.2, 2.1, 2.2, 3.1, 3.2, 4.1 and 4.2) consists of four images of the same word, phrase or clause. The top image illustrates its waveform; the bottom two its intensity on the left and its fundamental frequency (f_0) on the right, which is intended to be a pitch contour; and the middle one its formant frequencies – ‘formants’ are vocal tract resonances, whose images are shown as dark bands in a segment on a spectrogram. The individual distinctive pattern

of these dark bands is formed by the vocal tract's configuration of a speech sound. The lowest band is named as the first formant, the one with the next above the second formant and so on. Vowels, approximants and nasals are divided into segments with their own sets of formants in frequency.

Apart from the formants, there are vertical lines or striations in these sounds on the spectrograms. They indicate vibrations of the vocal cords. Another small vertical lines similar to the striations are making a dark or faint band along the base line, which is referred to as a voice bar. They show the vocal cords' vibrations of voiced consonants as well.

Mentioning the consonants in our spectrograms, the voiceless fricative [s] (cf [z]) produces a fuzzy segment that is darkest in the region of 4,000 to 6,000 cps (i.e. cycles per second); that of [ʃ] (cf [ʒ]) in the region of 2,000 to 3,000 cps. The voiceless plosives, [k] (cf [g]), [p] (cf [b]) and [t] (cf [d]), consist of three phases that are arranged in chronological order, that is, closure, shutting and opening phases. The last one depicts an explosion of high-pressured air having been blocked up in the oral cavity after the release of articulatory contact between the organs of speech. In our spectrograms, a blank segmented area of the plosive indicates its shutting phase, and the last phase illustrates a dark random pattern of the segment.

Examining acoustically, the fundamental frequency (f_0) – the vocal cords' frequency of vibration – in the male voice shows that 'founder' is slightly higher than 'conflict', and a falling contour is adopted by it. The image showing the intensity also proves that the level of the 'found-' is quite higher than those of the other syllables. Perceptually, too, the 'founder' is perceived bit higher in pitch than the 'conflict'.

The fundamental frequency (f_0) in the female voice is visibly rather constant, indicating that it is a level contour. The intensity image indicates that all the vowels in the 'founder conflict' are equally strong. Schematically it could be illustrated like a [●●●], so that it quite seems that these two nouns are standing independently of each other in the running speech rather than are getting together to compose a single compound noun.

6. Discussion

One of the major prosodic characteristics in English speech is its rhythm. Stressed syllables centering chiefly around content words are psychologically expected to tend to occur at regular intervals of time. When this phenomenon is observed in terms of external articulatory gesture, it shows various physiological movements involving articulation particularly on our face. It may also extend to our body as a whole. Stress is perceptually identified when any particular syllable is higher in pitch, longer in length, and/or louder in loudness. Vowel quality may also contribute to it because vowels are primarily the nucleus of a syllable and are more sonorous than consonants on the sonority scale. Any of these physiological efforts is achieved by greater muscle contraction.

Confining ourselves to the movement relating to the organs of speech, we as a listener could externally identify signs of (sentence) stress at least by means of the vertical motion of the jaw. Our two American subjects demonstrate this. In our experiment on reading-aloud in front of a camera and a teleprompter, the only source for information on (sentence) stress is externally their jaw movement. The following definition of stress by Laver (1994, 511) may strongly, although indirectly, support the close relationship between stress and the jaw motion,

Other things being equal, one syllable is more prominent than another to the extent that its constituent segments display higher pitch, greater loudness, longer duration or greater articulatory excursion from the neutral disposition of the vocal tract.

The vocal tract located at the air passages above the larynx and consists of the oral and the nasal tract within the nose. The former tract principally consists of the lips, teeth, hard and soft palates, uvula, tongue, and maybe the back wall of the pharynx (Ladefoged 1993). It does not include the jaw, but, in fact, necessarily involves in the lip and tongue movement. The lower jaw itself plays a major role of articulation when such organs of speech as the lips and tongue move to shape

itself for a particular speech sound.

Mostly, stress can be detected by perceptual pitch, loudness, and/or length, but in some cases this information does not satisfactorily give us the placement of stress (*or* segments). In everyday situation, our linguistic – phonological, syntactic, semantic – knowledge helps us as a listener to fill the missing perceptual information and understand the running speech.

The following examples will support the idea that we can perceive effectively and easily speech in various circumstances with our linguistic knowledge compensating for distorted speech sounds. The sentences below are taken from *The Speech Chain* (Denes and Pinson 1993). All we have to do is to substitute vowels for the asterisks. As for the last example, some consonants are also missing:

- (1) M*R H*D * L*TTL* L*MB H*R FL*C* W*S WH*T* *S SN *W
 (2) TH* S*N S* N*T SH*N*NG T*D**
 (3) S*M* W**DS *R* EA*I*R T* U*D*R*T*N* T*A* *T*E*S (181)

Similarly, the speaker will construct speech with the use of their linguistic knowledge. The surest thing to do is that they put (sentence) stress on certain words in every chunk of it to build up rhythm in English. The way they realize the stress in speech may vary internally. However, the essential external articulatory gesture should require jaw movement as our subjects have demonstrated in our reading-aloud experiment unless it is related with ventriloquism.

7. Concluding remarks

As a rule, the jaw is not included in the organs of speech, and, in a way, it is subsidiary to them. Thus, the physiological aspects of jaw motion seem to be treated like a by-product in speech analysis. The examination of external articulatory gesture in our experiment actually contends that the degree of departure from the neutral position of the jaw differentiates sentence stress from word stress. If a wider jaw motion purports the former stress in speech for the speaker, it could play as an alternative to its physical realization of rhythmic beats coinciding with syllabic nuclei.

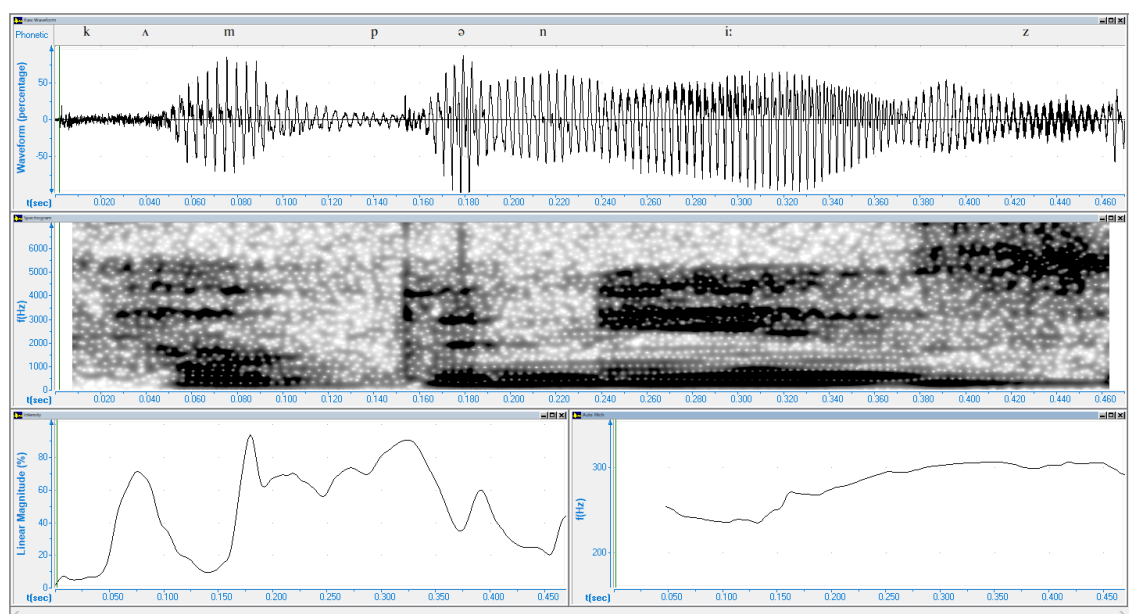


Figure 1.1 The waveform, spectrogram, intensity and f_0 of the word, ‘companies’, spoken by the female subject.

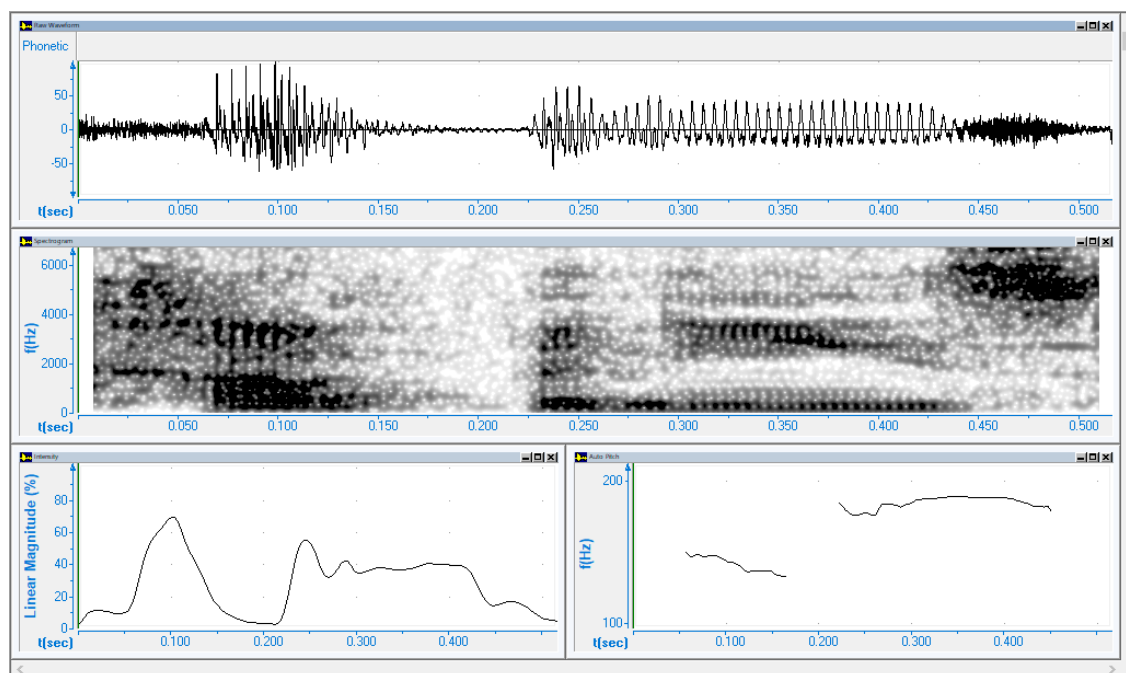


Figure 1.2 The waveform, spectrogram, intensity and f_0 of the word, ‘companies’, spoken by the male subject.

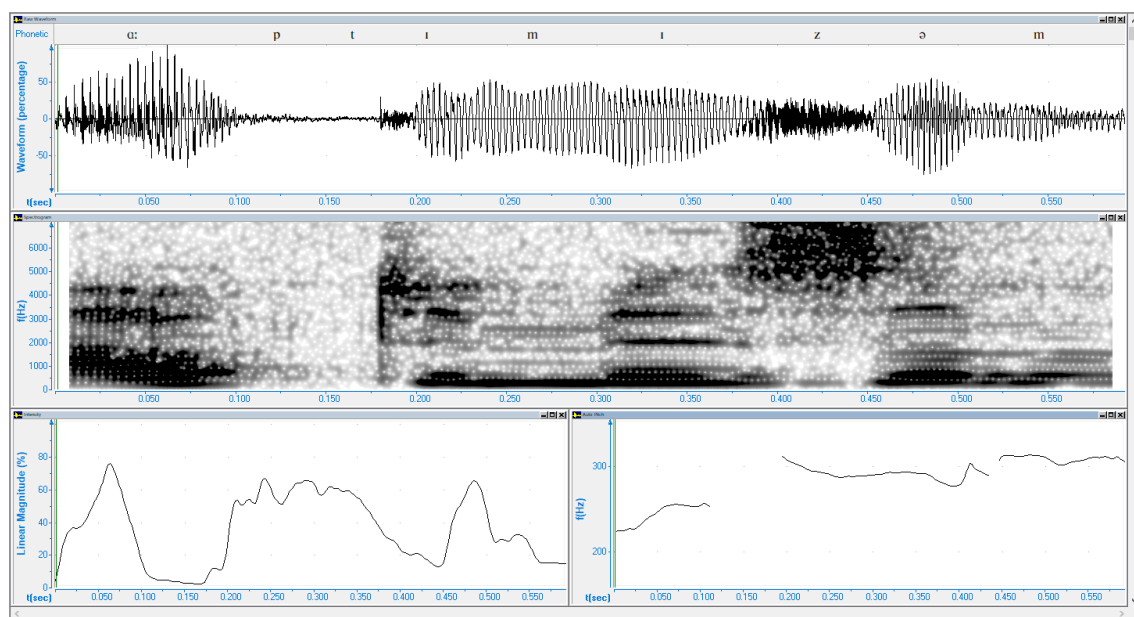


Figure 2.1 The waveform, spectrogram, intensity and f_0 of the word, 'optimism', spoken by the female subject.

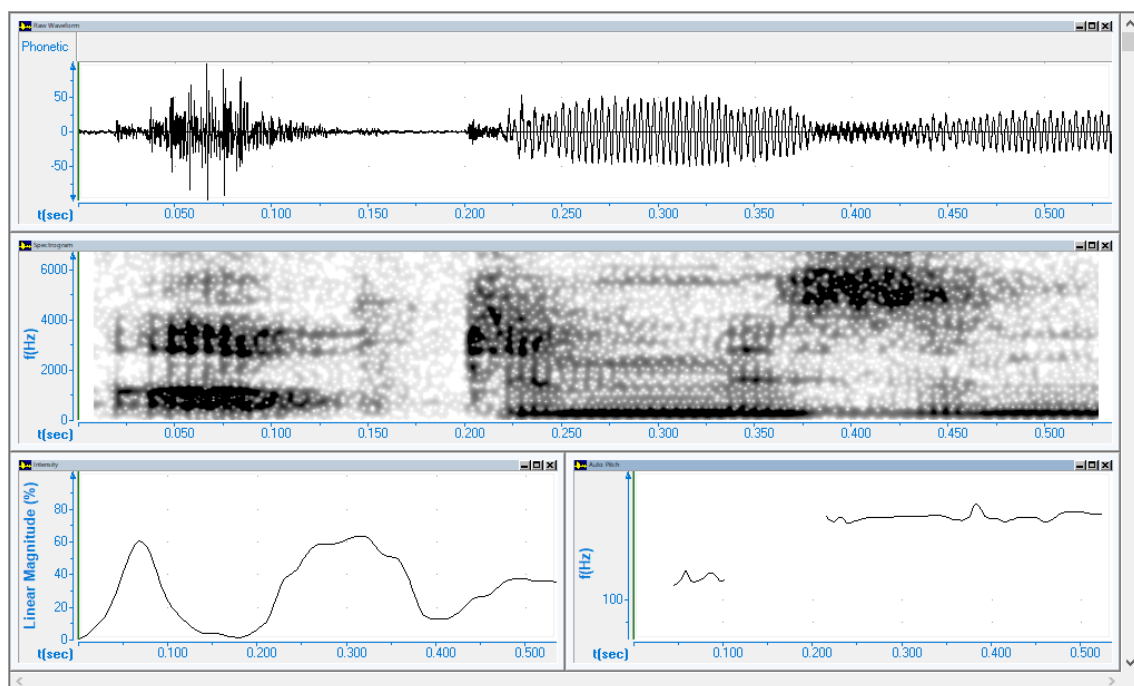


Figure 2.2 The waveform, spectrogram, intensity and f_0 of the word, 'optimism', spoken by the male subject.

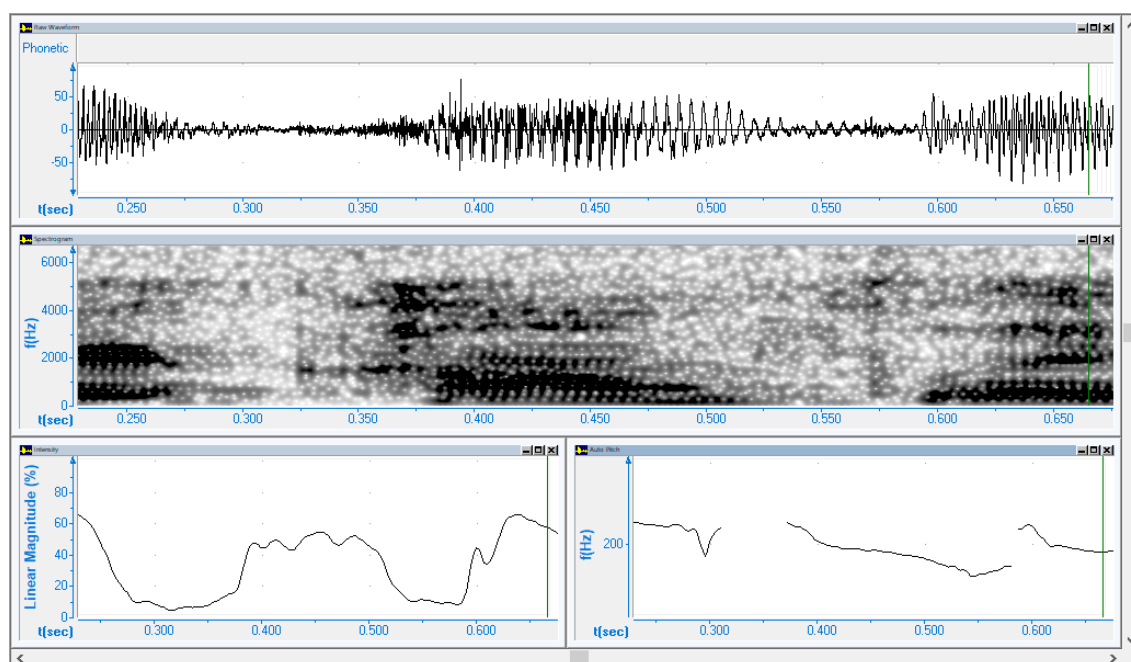


Figure 3.1 The waveform, spectrogram, intensity and f_0 of the phrase, 'founder conflict', spoken by the female subject.

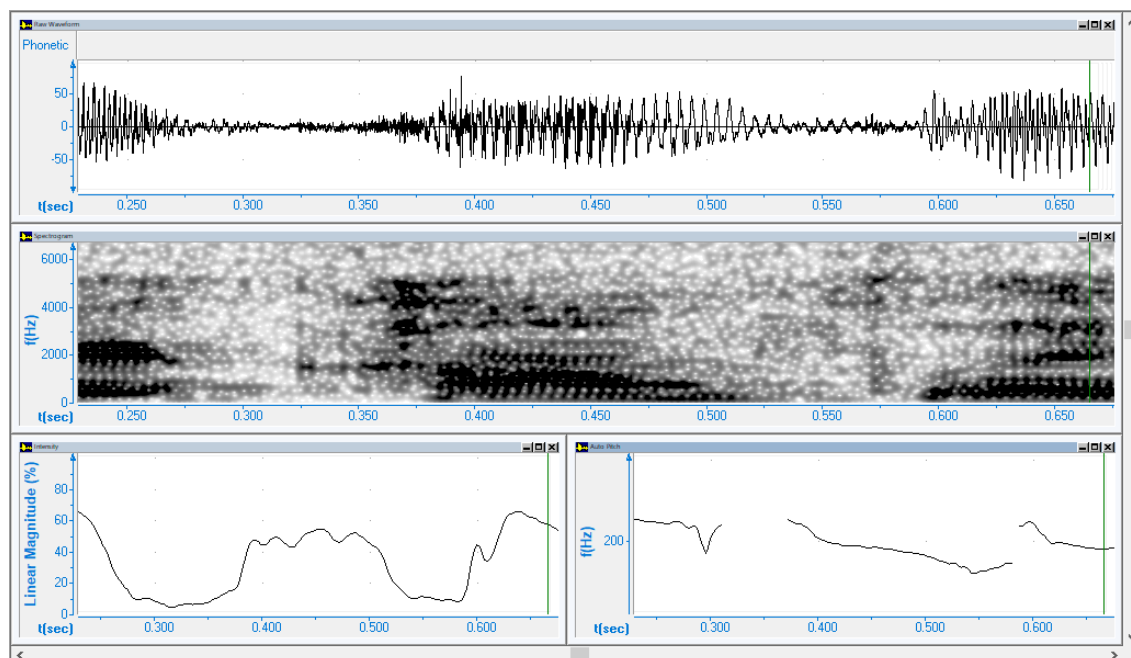


Figure 3.2 The waveform, spectrogram, intensity and f_0 of the phrase, 'founder conflict', spoken by the male subject.

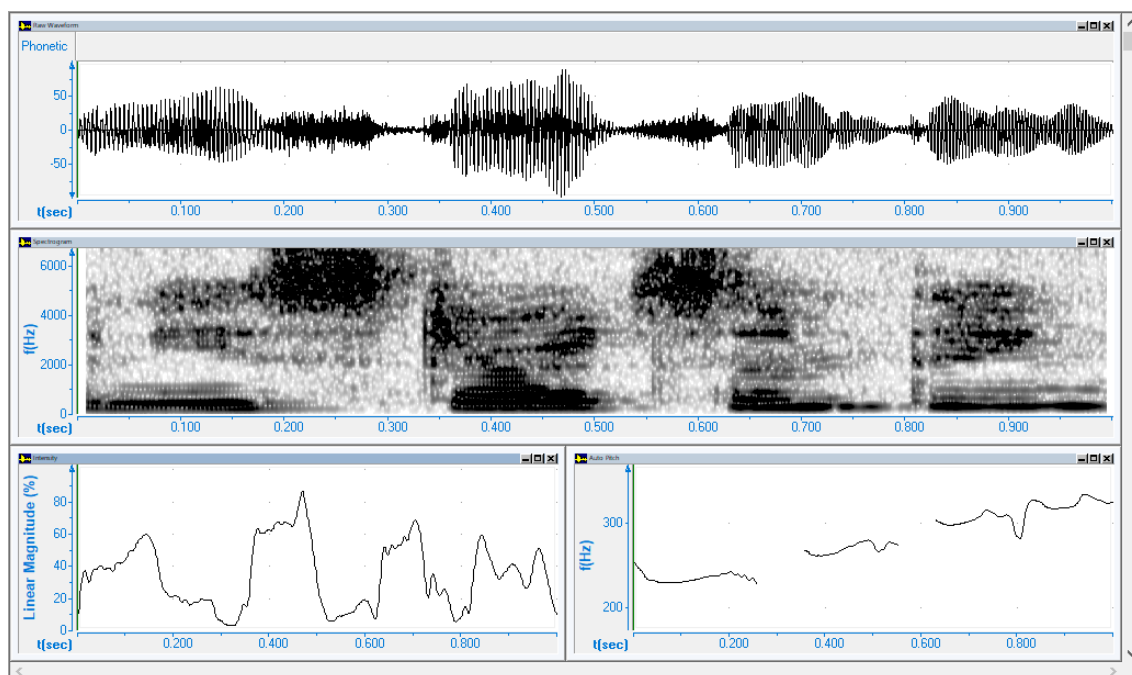


Figure 4.1 The waveform, spectrogram, intensity and f_0 of the clause, ‘when you start something’, spoken by the female

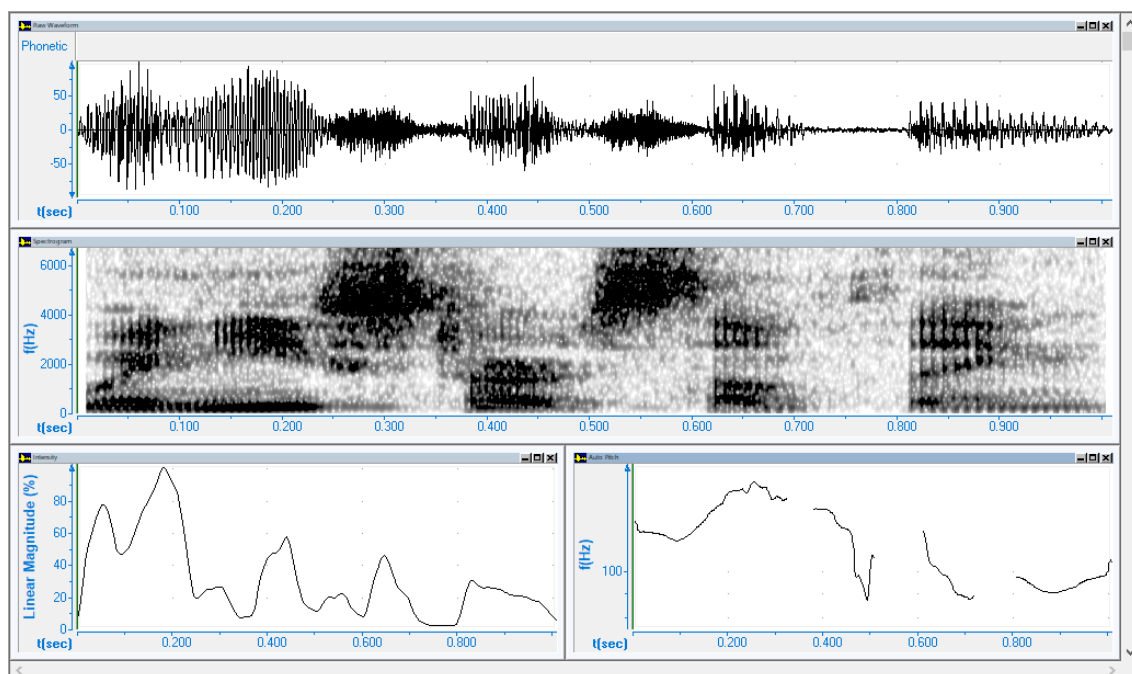


Figure 4.2 The waveform, spectrogram, intensity and f_0 of the clause, ‘when you start something’, spoken by the male

Part 3: Vertical Jaw Movement and Rhythm in Speech

1. Vertical jaw movement

In this last section, we concentrate on the pronunciation and scrutinize perceptually that of all the Japanese subjects. The main aim is to select those who fulfill the criteria below in order to compare them with the two American subjects in terms of jaw movement. This is the only external articulatory gesture that is most commonly identified in both subjects.

Our set of pronunciation criteria for the choice is as follows: the vowels and consonants, the intonation, the pause marking the phrase boundary, and the rhythm. The last criterion contains rhythmic timing, weak forms, contractions, consonant clusters and linking-ups.

As a result of the detailed examination, the following have been chosen:

- 7. RM
- 8. GT
- 26. YY (the head movement and articulation rate)
- 28. SW
- 31. TC (the facial expression quite the same as those two American students)
- 36. RT

A big difference between the two groups of subjects is the style of speech that they adopt in each experiment. The Japanese are told to memorize one of the two passages that they have been assigned to read aloud beforehand, whereas the Americans read one after another those two separate passages on the screen of a monitor placed in front of them. The difference between the groups will be quite distinctly found out in the aspect of articulatory gesture if the facial images on video are examined. It is visibly clear when we compare those subjects above except Subject 31 with the Americans. He is quite poor in overall external articulatory gesture, which is remarkably similar particularly to the American male subject.

The observation of our American subjects on video may imply that their external articulatory gesture is shared with native speakers of English in general. It seems reasonable to assume that this will be true, as Honikman (1964) comments articulatory settings about British English speakers, though, when engaging in everyday conversation,

In natural colloquial English, the jaws are, for the most part, held loosely together but not clenched — no tension is felt; the most frequent vowels appear to be [ɪ] and [ə] which do not require the jaws to be parted. There is from time to time some lowering of the jaw, but relatively infrequent and slight, so that the aperture between the upper and lower teeth is generally never wide — at most about a finger's width, as required for the diphthongs [aɪ, aʊ], less for [æ]. Thus it appears that the greater part of English articulation takes place behind (loosely) closed jaws. It is this feature of English, no doubt, which helps to give foreigners the impression that we do not move or open our mouths when we speak. (80)

2. The hearer as a speaker

With reference to rhythm in speech communication between native speakers, the speaker intends to produce an utterance that may not be acoustically isochronous but that may be perceptually and empathetically accepted as an isochronous speech product by the hearer. This could be predicted by the results of the experiments carried out by Lehiste (1973), Darwin and Donovan (1980), and Morton, Marcus and Frankish (1976), for example.

As to isochrony in English, Lehiste (1973) points out after carrying out the experiments on isochrony in production and in perception that perception may play a better role in spoken English. Focusing on a perceptual approach on rhythm in English, Darwin and Donovan (1980), as their results of their three experiments, they conclude that 'within a tone group, the perceived rhythm of speech is more rhythmically regular than is the speech itself'. In the world of psychology, Morton,

Marcus and Frankish (1976) present their subjects two of the spoken digits out of 1 through 9 with the one occurring at fixed intervals. They then instruct each of them to adjust a knob so as to make the other perceptually arranged at regular intervals of time. They prove that this psychological point in time (a ‘perceptual centre’) demonstrates that the acoustic onsets of the digits are not aligned regularly although their perceptual ones are doing so.

Exactly like the close connection between the speaker and the hearer, we ourselves are said to be simultaneously the speaker and hearer in our native tongue. This can be substantiated by the fact that we employ our feedback system when we deliver speech sounds. Feedback itself is not claimed to be essential for speech (Raphael, et al 2007), but we cannot do without it. It is also said to be crucial for the baby to acquire their own speech.

By and large, we as a speaker use four kinds of information that could be used in feedback control: auditory, tactile, proprioceptive, and central neural feedback (Raphael, et al 2007, 168) Auditory feedback is information about our own speech heard through our ears; tactile one is the information referring to, for example, the contact between the active and passive articulators during speech; proprioceptive one is information on muscles of articulators; and central neural one is the information concerning motor commands given by the brain, (however, ‘There is as yet no direct evidence for internal feedback.’ (174)).

Quite a famous experimental way for the speaker to notice that they themselves are the hearer is the ‘delayed speech feedback effect’ (Denes & Pinson 1993) (or ‘delayed auditory feedback (DAF) effect’ (Raphael, et al 2007) This is the effect that causes the speaker to start stammering or slurring when his own speech is fed back delayed by a fraction of a millisecond through headphones.

Apart from the experiment above, there are commonplace experiences we have from time to time. When we are talking with our friend on the underground train, for example, the volume of our voice dramatically increases once the train starts leaving the station into the tunnel. Incremental hearing loss also causes us to make our voice louder and, at the same time, to ask others to raise their voices.

3. Conclusion and Suggestions

Anatomically speaking, the vertical jaw movement may naturally happen to the accompaniment of the production of speech sounds. Comments about this by Laver (1980) may be worthwhile, as follows,

In continuous speech, most speakers visibly move their lower jaw vertically upwards and downwards, moving slightly for almost every segment. This perhaps is to be anticipated, given the intimate anatomical linkage ... between the jaw and the tongue and hyoid bone. This linkage is reflected in the way that the jaw can be seen to move in sympathy with the articulations of the body of the tongue. (63)

Adding in the above remarks, vertically lowering the jaw movement is slightly wider when it comes to a sentence-stressed syllable. This movement is made commonly by many of our subjects whether or not they are native speakers of English. At least, wider jaw movements as an external articulatory gesture seem to indicate that they coincide perfectly with rhythmic beats in speech.

Myth has it that Japanese learners of English hardly open their mouths and maintain a somewhat blank expression on their face when they speak. As far as our Japanese subjects in our experiment are concerned, the opposite might be the case with comparisons between them and those American subjects in the related experiment.

The myth in this case refers to the idea that many people believe, but which is not true. It is treated here exactly as the same meaning as *Language Myths* edited by Laurie Bauer and Peter Trudgill (1998),

As linguists, we are very much aware that ordinary people have well-established ideas about language. We meet these ideas when non-linguists talk to us at parties, in the common rooms of universities, from members

of our families and in the media. Some of these ideas are so well established that we might say they were part of our culture. It is in this sense that we refer to them as myths.... (From *Introduction*)

Our comprehensive assessment of Sections 1 and 2 directs our attention to the follow suggestion, The relationship between the lower jaw and sentence stress could be valuable information about rhythm in English for the foreign learners. They should therefore become conscious of stress at word level in English. Stress in English indicates the nucleus of the syllable, which corresponds to the stressed vowel. Each vowel has its own highest point of the tongue in the oral cavity with its own lip posture. The stressed one automatically moves the jaw downward further than the unstressed one as well as the consonant. Bearing this in mind, they should practice the pronunciation of various words in English. Then they should exercise actively lowering the jaw at sentence stresses to acquire the stress-timing. Without stress-timed rhythm, they could not give native speakers of English properly what they are saying.

References

- Abercrombie, A (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University.
- Bauer, L & Trudgill, P (1998). *Language Myths*. London: Penguin Books.
- Cakewalk (2008). *Audio Creator LE(x64)*. Cakewalk Music Software.
- Darwin, C J & Donovan, A (1980). Perceptual studies of speech rhythm: isochrony and intonation. In J C Simon (ed), *Spoken Language Generation and Understanding*, D Reidel Publishing Company, 77-85.
- Denes, P B & Pinson, E N (1993). *The Speech Chain: The Physics and Biology of Spoken Language*. New York: W H Freeman and Company.
- DigiOn (2016). *DigiOnSound X (version 1.1.35)*. DigiOn, Inc.
- Honikman, B (1964). Articulatory settings. In D Abercrombie, D B Fry, P A D MacCarthy, N C Scott & Trim, J L M (eds), *In Honour of Daniel Jones*, 73-84. London: Longman.
- Ladefoged, P (1993). *A Course in Phonetics (3rd ed)*. Fort Worth: Harcourt Brace College Publishers.
- Laver, J (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
- Laver, J (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lehiste, I (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America* 54, 1228-1234.
- Masters, B (2014). *Zero to One (unabridged)*. New York: Random House Audio.
- Morton, J, S Marcus & Frankish, C. (1976). THEORETICAL NOTE: Perceptual Centers (P-centers), *Psychological Review* 83 (5), 405-408.
- Parker, S (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics* 36, 55-90.
- Raphael, L J, G J Borden & Harris, K S (2007). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech (5th ed)*. Philadelphia: Lippincott Williams & Wilkins.
- SL International (2012). *Speech Analyzer (version 3.1)*. <https://software.sil.org/speech-analyzer>.
- Tamaki, T (1998). Speech rhythm autonomy. *Ronshu* 39, 29-45. Tokyo: Aoyama Gakuin University.
- Thiel, P with Masters, B (2014). *Zero to One*. New York: Crown Business.

Appendix

‘Companies ^ are like \countries in ,this ,way.// ‘Bad \decisions ,made ,early ,on //– if you ‘choose the ‘wrong \partners ^ or ‘hire the ‘wrong \people, for e\example //– are ‘very ‘hard to cor\rect ^ after they are \made. It may ‘take a \crisis on the ,order of ,bankruptcy ^ before ‘anybody will ‘even ‘try to cor\rect them.// As a /founder,/ your ‘first \job ^ is to ‘get the ‘first ‘things \right, because you \cannot ,build a ,great ,company/ on a ‘flawed foun\da-tion.//

When you \start ,something,// the /first ^ and 'most 'crucial de\cision you ,make ^ is 'whom to \start it with.//
 'Choosing a co-/founder ^ is like 'getting \married,// and 'founder \conflict ^ is 'just as \ugly as di,vorce.// 'Optimism
 a\bounds at the ,start of ,every re,lationship.// It's unro'mantic to 'think \soberly ^ about 'what could 'go \wrong,// so
 'people \don't.// But if the /founders ^ de'velop irrecon'cilable /differences,// the 'company be'comes the \victim.//

//: long pause; /: short pause; ^: tone group boundary; ' high-level sentence stress; ,low-level sentence stress; \falling nucleus; \low-falling nucleus; /rising nucleus; /low-rising nucleus; words in bold letters are those with a tonic syllable in a tone group, which is delimited by pause. The tonic syllable (also called tonic, nucleus (*or* nuclear syllable)) carries the main movement of pitch. A tone group is alternatively called a 'phonological phrase' *or* an 'intonation group'.

The above two passages taken from *Zero to One* (Thiel, et al 2014) are perceptually analysed based on the original recordings (available from Audible Audiobooks, Random House Audio) made by one of the co-authors, Blake Master