Formant movement of labial-velar and palatal glides in Japanese: Preliminary acoustic data with young adult female speakers

Makoto KARIYASU

Abstract

To present a method and to obtain preliminary normative data of formant movement of labial-velar and palatal glides in Japanese, each of 11 speakers produced a series of /awa/ and /aja/ as well as two sentences containing these VCV tokens. An acoustic analysis using a wide-band spectrogram and LPC formant tracking was performed to measure durations of VCV and of the segment of formant changes, and the first formant (F1) and the second formant (F2) at the beginning of on-glide and the end of off-glide and the mid-point (center) of consonant. The rate of F1 and F2 movement (kHz per second) was then computed. The results showed that the patterns of F1 and F2 movements for /awa/ and /aja/ were mostly symmetric in time and consistent regardless of the speakers and speaking conditions. The rate of F1 movement for /w/ and for /j/ was similar, with a grand mean of 4.49 kHz/s and individual data ranged from 2.13 to 7.28kHz. The rate of F2 movement for /j/(a grand mean of 8.61 kHz/s) was greater than that for /w/ (a grand mean of 3.47 kHz/s). No systematic trend in the rate of formant movememnts between syllable and sentence productions although the speaking rate was faster (conversely, VCV duration was shorter) in sentence production than in syllable repetitions. These acoustic data, which are assumed to reflect normal articulatory gesture of the adult speakers, will be used as preliminary normative data when speech data of aged adults and patients with speech disorders are evaluated.

Key words : Glides, Japanese, Acoustic analysis, Formant movement, Normal speech production

Introduction

Principles of Speech Acoustics and Clinical Applications

Speech signal is a product of laryngeal voice source (and/or vocal tract noise) and the vocal tract filtering (i.e., resonance). This source-filter theory of speech production is widely accepted, and one of the scientific accounts is to relate vocal tract configuration to acoustic properties (i.e., resonance frequencies, namely formants, observed in speech spectra) of speech signal [1]. The relationship between acoustic speech output and vocal tract configuration is fairly predictable; that is, acoustic properties of speech sounds reflect articulatory gestures or oral and laryngeal movements for the sound production [2].

Formants are analogous to the vocal tract resonances, and are shown as prominent (peak) frequency regions in a speech spectrum or as dark bands in a sound spectrogram. Several principles regarding formants are listed here[2-4]. First, the formants reflect the vocal tract length; the longer the vocal tract, the lower the formants. Second, the first formant (F1) reflects the mouth opening; closing the mouth results in F1 lowering. Third, the second formant (F2) is affected by the tongue advancement ; the tongue body placed at front in the oral cavity makes

九州保健福祉大学　保健科学部　言語聴覚療法学科　〒882-8508　宮崎県延岡市吉野町1714番地1号
Department of speech therapy, School of health sciences, Kyusyu University of Health and Welfare
1714-1 Yoshino-cho, Nobeoka, Miyazaki 882-8508 JAPAN　E-mail: kariyasu@phoenix.ac.jp

higher F2, and F2 decreases as the tongue body retracts. These principles or rules are quite useful when the acoustic properties are related to one's vocal tract size and oral movements for speech production.

The source-filter theory and the principles of speech acoustics have a potential for the description of the articulatory gestures in both normal and abnormal speech production[1,3,4]. Normal speech production in developmental and aging processes has been studied with phonetic and acoustic descriptions. Because the acoustic analysis is a non-invasive technique, it will be used extensively in speech clinics for the evaluation and treatment of speech disorders. When abnormal speech articulation (i.e., oral movements) is suspected, consonants and vowels can be selected and analyzed. One class of the speech sounds that involves oral movements is glides. The following section provides an overview of the production and acoustic properties of glides.

Production of Glides and their Acoustic Characteristics

Glides are a class of consonants produced with a laryngeal voice source and dynamic oral constriction. The oral constriction is mild so that noise source will not be generated along the vocal tract like vowels and nasals. They are also called soronant consonant, semi-vowels, or approximants. Two glide consonants, /w/ and /j/, are found in Japanese as well as in English. The place of articulation is labial (labial-velar) for /w/ and palatal for /j/. The glides are produced with greater oral constriction than vowels with similar vocal tract configurations : A labial-velar glide /w/ has a similar vocal tract configuration to the high-back vowel /u/; a palatal glide /j/ is similar in the vocal tract configuration to the high-front vowel /i/. They are produced with a gradual oral movement from and to an adjacent vocalic segment such as a vowel. Therefore, the glides when combined with vowels are similar to vowel combinations, diphthongs (i.e., /wi/-/ui/, /ju/-/iu/)[3] ; the differences are that the glides are produced with a greater constriction and that the articulatory movements for glides are faster than those for diphthongs [5]. When the glides are produced in intervocalic position (i.e., a consonant placed between vowels), the relevant articulatory adjustments for glides /w/ and /j/, such as tongue body position, lip protrusion only for /w/, tongue root advancing, must be accomplished from the positions for adjacent vowels; the total time taken to perform these maneuvers is in the range of 200 to 250 ms [6]. Their acoustic characteristics are lower formant frequencies and their movement [3-6]. Both labial-velar and palatal glides are associated with F1 lowering. The F1 for the glides estimated and measured is between 250 and 300 Hz in adult male and female speakers though technical difficulty in the measurement of F1 exists due to the overlapped fundamental frequency [5]. However, differences in the place of articulation (i.e. vocal tract constriction) associated with these glides result in very different spectral shapes or formant movememnts above F1. Due to the constriction at labial and velar regions, F2 is further lowered for /w/. The F2 is estimated about 700 Hz, below 1000 Hz. A weak prominence of F2 adjacent to the dominant peak of F1 can be seen. An amplitude of F3 decreases because of low F1 and F2. When a labial-velar glide /w/ is produced in intervocalic position, both F1 and F2 tend to move downward into the glide and upward following the release of the glide. In contrast, the palatal glide /j/ produced with the tongue blade against the palate gives rise to F2 that is high and is close to F3 and F4. The F2 is estimated greater than 2000 Hz. Both F2 and F3 reach a maximum at the mid point of glide.

Purposes of the Present Study

Acoustic data are thought to be useful means to quantify the characteristics and abnormality of speech production (i.e., articulatory movement). The glides in intervocalic position can be easily produced in children and adults with speech disorders and are ideal for estimating oral movements for speech production. Because the sources of normative data for glides in Japanese are limited and the acoustic data are in part dependent on speech material and sampling-analysis routine, the normative data with a standard method are needed. Therefore, the purpose of this study is to describe a standard sampling and analysis routine and to obtain preliminary normative data for the formant movement of glides in Japanese-speaking young adults.

Method

Participants

Eleven young female adults with age ranged from 19 to 22 years old participated in the present investigation. All reported no history of hearing and speech deficits. Their physical conditions were fair to good on the day of data collection. Though the speakers differ in dialect, it is not considered to have a substantial impact on the formant movement in syllable and connected speech productions.

Material

Labial-velar and palatal glides /w/ and /j/ were embedded in a low mid vowel /a/ in a vowel-consonant-vowel (VCV) context. The VCVs, /awa/ and /aja/, were also included in one of the short sentences at a word-initial position. The following is a list of two sentences in a broad phonetic transcription and in Japanese.

(1) [sonoutawa awatembo:no santakuro:su desu]
その歌は､『あわてんぼうのサンタクロース』です。

(2) [sono tot∫itotatemonowa aja∫i: buʔkendesu]
その土地と建物は、あやしい物件です。

Procedure

Each speaker was instructed to produce /awa/ and /awa/ five times with her comfortable voice and used tempo. Then, two sentences were produced in the same manner with a single breath group. All the productions were recorded with a microphone (BDM-100, Bose) coupled with a digital tape deck (TCD-D10, Sony).

Analysis (Figure 1)

Recorded speech signal was digitized at 22 kHz and was analyzed with an acoustic analysis software (Multi-Speech 3700, Key Elemetrics) housed in a personal computer (8200, Dell). Middle three VCV repetitions, excluding initial and final productions, and one VCV in a sentence for each glide were analyzed. First, a target segment (i.e., /awa/ or /aja/) was identified, and total duration of the target VCV was measured based on the speech waveform and spectrographic displays. Then time at which F1 and F2 start to move (on-glide) and time at which F1 and F2 terminate to move and reach steady-state vowel F-pattern (off-glide) were visually determined. The duration between on-glide and off-glide points was recorded. At these two points, F1 and F2 were also measured based on a wide-band spectrogram and LPC (Linear Predictive Coding) formant tracking. Then, F1 and F2 at the middle of glide (consonant), where a maximum of F1 and F2 excursion were observed, were recorded. Finally, the rate of F1 and F2 movement (kHz per second) was computed. Statistical software (SPSS) was used for the computation of summary statistics and correlation analysis.

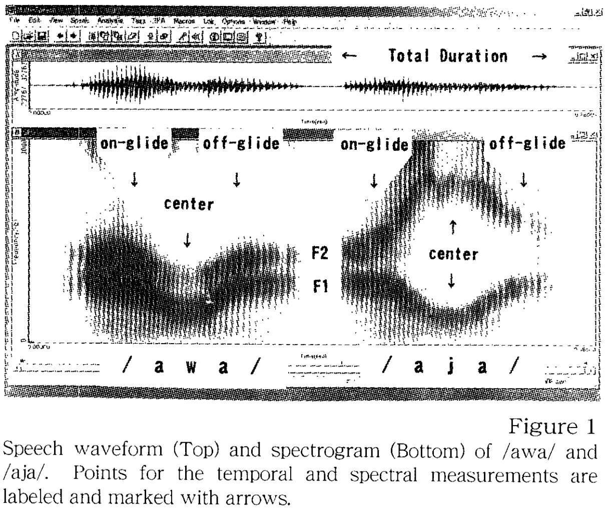

Figure 1
Speech waveform (Top) and spectrogram (Bottom) of /awa/ and /aja/. Points for the temporal and spectral measurements are labeled and marked with arrows.

Reliability

Selected samples were reanalyzed by the investigator a few weeks after the initial analysis. Two VCVs and one VCV in a sentence from each speaker were randomly chosen. The differences in the rate of formant movement between the first and second measurements were less than 20%.

Results

Temporal and spectral characteristics of labial-velar and palatal glides in a VCV context, /awa/ and /aja/, during syllable repetition and during sentence production are shown in Table 1 and Table 2, respectively.

Total duration of /awa/ and /aja/ was 343.5ms (262.7-427.8ms) for /awa/ and 322.9ms (273.1-398.3ms) with considerable individual variations (slightly greater than 10 % of the coefficient of variations). The total duration was reduced by 59% (45-71%) for /awa/ and 66% (45-77%) for /aja/ during sentence production relative to that during syllable repetitions. Predicted patterns of the formant movememnt were observed in all the speakers. Both F1 and F2 decreased for labial-velar glide /w/; F1 lowered and F2 raised for palatal glide /j/. These formant changes were symmetric in time (see Figure 1). The formant patterns held constant both during syllable repetitions and during sentence production.

It was uniform across the speakers, target VCVs, and speaking conditions, that F1 was relatively high (695-1097 Hz) for the vowel /a/. F2 was relatively low and close to F1 for /awa/ but it is considerably higher for /aja/. The rate of F1 movement for /w/ and for /j/ was similar, with a grand mean of 4.49kHz/s and individual data (speaker means) ranged from 2.13 to 7.28kHz/s. The rate of F2 movement for /j/ was greater with means of 8.40kHz/s for on-glide and 8.81kHz/s for off-glide. This greater rate of F2 movement for /aja/ was accounted for by higher F2 observed for /aja/. No uniform trend in the rate of F1 and F2 movement was found when comparing syllable repetitions with sentence productions. Correlational analyses were performed; however, any systematic relationship between the total duration and the rate of F1 and F2 movements was not observed.

Discussion

The patterns of formant movements for labial-velar and palatal glides in intervocalic position, /awa/ and /aja/, were uniform and consistent with previously reported data and predictions[1,3-6]. F1 falls toward the glide production for both /w/ and /j/. F2 descends for /w/ whereas it elevates for /j/. The rate of F1 movement was fairly invariant for /w/ and /j/. The rate of F2 movement for /aja/ is greater than the rate of F1 movement.

Relationship between Vocal Tract Configuration and Acoustic Properties

The patterns of formant movement observed in glide productions are reflections of oral articulatory gestures. As discussed in the earlier section, a close relationship between vocal tract configuration and acoustic properties is expected like other classes of the speech sounds. Occluding the mouth for /w/ in addition to lip protrusion makes F1 lowering; labial and velar constriction for /w/ also results in low F2. Similarly, the mouth is closing for /j/; as a result, F1 for /j/ tends to be low like for /w/. In contrast, advanced tongue body in the oral cavity for /j/ causes F2 elevation. It is inevitable that these patterns of formant movements for glides in intervocalic position are found for normal adult speakers in the present study.

The uniform and consistent pattern of formant movements is used as a qualitative index of normal speech production. Although the magnitude of formant changes and the duration, hence the rate of F1 and F2 movements, somewhat vary among adult speakers, the rate of F1 and F2 movement can be used as a quantitaive index of normal articulatory gestures.

A Standard Method

The speech material used in this study was /awa/ and /aja/ and they are sampled in two speaking conditions. Though the speaking rate derived from the total duration of VCV segment was slower in syllable repetition than in sentence production, the patterns and the rate of formant movements remain indifferent.

An acoustical analysis of the central frequency (prominence or peak) of resonance is not always easy and not free from measurement errors. For the measurement of F1 for glides, acoustic energy (i.e., voice bar in spectrogram at the low frequency region) due to the fundamental frequency of voice source may overlap the band of F1. In the present study, F1 values were higher than estimated values and measured data provided by Stevens [6]. On the other hand, measuring F2 for /aja/ was fairly straightforward, and the measurements seemed highly replicable. The judgement of on-glide and off-glide and the maximum excursion for the glides is another methodlogical issue. It is known that the articulatory adjustment takes approximately 100 ms from a vowel to a glide and a glide to a vowel. Formant movement duration data in this study was around 150 ms in syllable repetition and slightly greter than 100 ms in sentence production. It seems relevant to the previous data and expectations.

Continuous efforts to search for speech material and acoustic indice are warranted. The formant was measured by eye balls and hands in this study, and it may introduce some errors. It is ideal to develop a system with automated segment identification and formant measurement.

Research and Clinical Utilities

Traditionally the assessment of speech or articulation involves perceptual judgement of skilled clinicians or phoneticians. An acoustical analysis will support the

Table 1. Temporal and spectral characteristics of labial-velar and palatal glides in a VCV context, /awa/ and /aja/. Mean, standard deviation, and range of each measurement from Japanese-speaking 11 adult female speakers are presented.

	/awa/		/aja/	
Total duration (TD, ms)	343.5 (53.7)	262.7～ 427.8	322.9 (35.9)	273.1～ 398.3
Formant movement duration (FMD, ms)	149.7 (24.1)	106.6～ 187.5	169.4 (17.4)	144.2～ 188.2
Vowel F1 at the on-glide (Hz)	905.7 (95.0)	783.3～1092.3	907.8 (100.9)	784.7～1097.0
Vowel F2 at the on-glide (Hz)	1435.8 (85.4)	1288.0～1563.3	1704.4 (99.4)	1532.7～1834.3
Consonant F1 (Hz)	547.0 (48.3)	458.7～ 644.0	523.5 (45.3)	468.0～ 595.7
Consonant F2 (Hz)	1173.2 (80.8)	1017.0～1307.0	2408.8 (170.2)	2028.7～2742.3
Vowel F1 at the off-glide (Hz)	868.5 (55.6)	789.3～ 964.3	871.3 (67.4)	789.3～ 983.3
Vowel F2 at the off-glide (Hz)	1417.8 (67.9)	1295.3～1546.0	1671.1 (65.6)	1569.3～1768.3
On-glide F1 Rate (kHz/s)*	4.90 (1.30)	2.72～ 7.28	4.55 (1.06)	3.07～ 6.33
On-glide F2 Rate (kHz/s)*	3.56 (0.62)	2.64～ 4.53	8.40 (2.01)	5.29～12.54
Off-glide F1 Rate (kHz/s)*	4.37 (0.87)	3.34～ 6.17	4.13 (0.90)	3.07～ 5.91
Off-glide F2 Rate (kHz/s)*	3.37 (1.05)	2.13～ 5.22	8.81 (2.01)	4.75～12.47

*The rate of Fn change for on-glide and off-glide is computed by the difference between vowel and consonant Fn devided by a half of formant movement duration. $\Delta Fn = |Fn\ (vowel) - Fn\ (consonant)| / (FMD \times 0.5)$

Table 2. Temporal and spectral characteristics of labial-velar and palatal glides /awa/ and /aja/ in a sentence, including proportions to the measures of VCV production (see Table 1) in the temporal domain. Mean, standard deviation, and range of each measure from Japanese-speaking 11 adult female speakers are presented.

	/awa/		/aja/	
Total duration (TD, ms)	200.6 (17.9)	159.2～216.8	214.2 (28.1)	161.7～250.1
Formant movement duration (FMD, ms)	104.3 (14.0)	80.7～126.8	123.5 (16.8)	91.0～143.4
Proportion of TD to TD in VCV	0.59 (0.09)	0.45～ 0.71	0.66 (0.09)	0.45～ 0.77
Proportion of FMD to FMD in VCV	0.71 (0.13)	0.52～ 0.85	0.73 (0.09)	0.57～ 0.85
Vowel F1 at the on-glide (Hz)	843.3 (71.4)	723～ 965	840.8 (66.5)	695～ 950
Vowel F2 at the on-glide (Hz)	1413.1 (96.8)	1248～1574	1716.4 (68.4)	1631～1830
Consonant F1 (Hz)	627.7 (91.3)	425～ 738	591.9 (67.3)	511～ 695
Consonant F2 (Hz)	1215.8 (71.3)	1078～1319	2232.0 (139.7)	2085～2567
Vowel F1 at the off-glide (Hz)	826.6 (50.6)	738～ 908	794.3 (57.9)	695～ 908
Vowel F2 at the off-glide (Hz)	1502.0 (88.3)	1333～1602	1891.6 (113.5)	1730～2156
On-glide F1 Rate (kHz/s)	4.05 (0.94)	2.30～5.82	3.96 (0.96)	2.18～ 5.57
On-glide F2 Rate (kHz/s)	3.78 (1.21)	2.30～6.17	8.31 (1.80)	5.82～11.80
Off-glide F1 Rate (kHz/s)	3.76 (1.46)	1.84～6.06	3.25 (1.14)	2.07～ 5.05
Off-glide F2 Rate (kHz/s)	5.46 (1.75)	2.22～7.72	5.57 (1.63)	2.28～ 7.34

judgement with quantitative data[4]. The acoustic-phonetic description of speech is needed to illustrate individuals with or without speech disorders in different age ranges for various aspects of phonetics and speech-language pathology.

Formant movement, particularly F2 trajectories, was found to be an index of speech motor function. It was highly related to speech intelligibility in adult patients with a degenerative neuromuscular disease, ALS [7-9]. Currently, Bunton and Weismer (2002) provided some segment-level measurements that could be sensitive to changes of speech function, A protocol of acoustic description in addition to currently available means of speech testing would make a package of outcome measures and provide diagnostic implications for the speech production system.

For the evaluation of speech data in the motor speech disorders, i.e., dysarthrias, these normative data, though it is preliminary, are valuable. Speech in the aged persons is perceptually normal and defective articulation (i.e., sound production) implies any physiological (sensory, motor, and/or cognitive) disturbance. It is possible that subtle imperceptable changes in speech signal may be detected with an acoustic analysis. An acoustic-phonetic (perceptual) description of individual speakers is also possible, and it may reveal genetic and environmental influences on human speech. It is definitely required to accumulate data from older adults and children to establish the norm. With the data of development, maturation, and aging of speech production, our knowledge of the speech production system will be enhanced, which certainly contributes to research and clinical applications.

Acknowledgement

The author wishes to thank students who voluntarily served as research participants. This investigation was in part supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Exploratory Research No.13871063 (FY2001-2002) to the author.

References

1) Fant G: The Acoustic Theory of Speech Production. Mouton, 1960.

2) Lindblom BEF and Sundberg JEF: Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of Acoustical Society of America, 50, 1166-1179, 1971.

3) Pickett JM: The Acoustics of Speech Communication. University Park Press, 1980.

4) Kent RD and Read C: The Acoustic Analysis of Speech. Singular Publishing Group, 1992.

5) Pickett JM (ed.): The Acoustics of Speech Communication - Fundamentals, Speech Perception Theory, and Technology (Chapter 6). Allyne & Bacon, 1998.

6) Stevens KN: Acoustic Phonetics. MIT Press, 1999.

7) Kent RD, Sufit RL, Rosenbek JC et al.: Speech deterioration in amyotrophic lateral sclerosis - A case study. Journal of Speech and Hearing Research, 34, 1269-1275, 1991.

8) Kent JF, Kent RD, Rosenbek JC, et al.: Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. Journal of Speech and Hearing Research, 35, 723-733, 1992.

9) Mulligan M, Carpenter J, Riddel J et al.: Intellgibility and the acoustic chracteristics of speech in amyotrophic lateral sclerosis (ALS). Journal of Speech and Hearing Research, 37, 496-503, 1994.

10) Bunton K and Weismer G: Segmental level analysis of persons with motor speech disorders. Folia Phoniatrica et Logopaedica, 54, 223-239, 2002.

日本語の口唇－軟口蓋および口蓋わたり音のフォルマント移動：若年成人女性の初期音響データ

苅 安 誠

九州保健福祉大学 保健科学部 言語聴覚療法学科

抄 録

日本語の口唇－軟口蓋・口蓋わたり音のフォルマント移動の測定方法を示し、その標準値を求めるために、成人女性話者11名に標的音声／awa／と／aja／の連続とそれらを含む文を生成させた。広帯域スペクトログラムと線形予測法（LPC）によるフォルマント軌跡を用いた音響分析により、標的音声とフォルマント変化の持続時間、わたりの開始・終了地点と子音中央部の第１フォルマント（F1）と第2フォルマント（F2）を測定した。これらの測定値より、フォルマント変化率（kHz/s）を算定した。この結果、/awa/と/aja/のF1とF2の遷移パターンは、時間的に対称性をもち、話者および発話条件にかかわらず一貫していた。F1の変化率は/w/・/j/とも近似し、全体平均で4.49kHz/s、話者個人では2.13～7.28kHz/sであった。F2変化率は、/j/（全体平均8.61kHz/s）の方が/w/（全体平均3.47kHz/s）よりも大きかった。文の生成で話速度が高く（反対に、VCV持続時間は短く）なってはいたが、F1とF2の変化率に一貫した傾向を認めなかった。このような音響データは、話者の正常な調音運動を反映すると考えられ、高齢者や構音障害の患者の発語を評価する際に参照する標準データとして用いることになる。

キーワード： わたり音、日本語、音響分析、フォルマント移動、正常な音声生成