

An Acoustic Study of President Obama's Inaugural Address

Kazuaki ICHIZAKI

Associate Professor, Miyazaki Gakuen Jr. College, Japan

email: ichizaki@mwjc.ac.jp

要旨

オバマ大統領の就任演説を音声分析ソフトを用いて分析した。平均した発話速度は政治家の演説としては標準的で、使用するポーズの頻度も標準値に近かったが、テレプロンプターの活用もあって言い淀みはほとんどなく、使用ポーズのすべてが silent pause となった。他方、音調群の長さや音調群を構成する音節数には幅があり、それが伝達情報量の圧縮率にメリハリを生じさせていることが判明した。また、特に強調したい音調群では発話の速さは減じる傾向があり、その直前の音調群では発話の速さが大きくなった。更に、結びとなる節の直前には長いポーズを置き、その節の第一音調群では、それを構成する内容語で音圧が最大となった。大観衆を目前にしてもあくまでも会話風に語りつつ、ここぞという場面では、ポーズを効果的に使ったり、発話の速さを変化させたり、語気を強めるといふ、語りの妙が観察された。

1. Introduction

Although his popularity seems to be decreasing these days due to stagnant businesses or sluggish reformation of the medical insurance system, nobody would deny that Barack H. Obama, the President of the United States, is one of the most eloquent speakers among the forty-three presidents in her history. The impression of his speeches is so cool and he seldom shouts or gets excited, instead he controls himself well and pays much attention to his audience right and left.

It seems to be true that he is an expert user of the teleprompter as his eye attention is rarely paid to it. He always makes speeches just like he talks to his partners in his daily conversation. Apart from gross and fine motor skills, eye attention, and ad lib remarks, his speech techniques are going to be investigated from an acoustic point of view by focusing on his inaugural address done in 2009. It is the aim of this paper to show a model of ideal public speech with objective data. Amateur speakers could become somewhat eloquent by imitating and adopting the phonetic features of the model.

2. Investigation and Analysis

2.1 speech material

The inaugural address was done at noon on January 20th, 2009 in Washington D. C. with a very large audience, which was said to be as many as two million. The address lasted for 18.5 minutes, which was a rather short speech for an inaugural address. A piece of the address which he seemed to have highlighted was extracted for the sound material for this paper. It was separated into tone groups as follows for the convenience of making the tables listed at the end of this paper.

Extract :

That we are in the midst of crisis || p1 *is now well understood.* || P2 *Our nation is at war;* || P3 *against a far-reaching network* || P4 *of violence and hatred.* || P5 *Our economy is badly weakened, a consequence of* || P6 *greed and irresponsibility on the part of some,* || P7 *but also our collective failure to make hard choices* || P8 *and prepare the nation* || p9 *for a new age.* || P10 *Homes have been lost;* || p11 *jobs shed;* | *businesses shuttered.* || P12 *Our health care is too costly;* | *our schools fail too many;* || P13 *and each day brings further evidence that the ways we use energy strengthen* || P14 *our adversaries and threaten our planet.* || P15

These are the indicators of crisis, || P16 *subject to data and statistics.* || P17 *Less measurable* || P18 *but no less profound is a sapping of confidence* || P19 *across our land —* || P20 *a nagging fear that* || p21 *America's decline* || P22 *is inevitable, ~~and~~ that* || P23 *the next generation must lower its sights.* || P24

Today I say to you that the challenges we face are real. || P25 *They are serious and they are many.* || P26 *They will not mq be met easily* || P27 *or in ~~a~~ short span of time.* || P28 *But know this, America —* || p29 *they will be met.*

| : the end of a tone group without pause

|| : the end of a tone group with pause

pNo.: a pause without a breather

PNo.: a pause with a breather

~~be~~ : word not pronounced

~~ : wrong utterance

2.2 procedure

The sound material was extracted from “You tube” items titled ‘Barack Obama Inaugural Address¹⁾’ which originally seemed to have been broadcast through “Sky News,” on the internet. The materials, which were played by Adobe flash player, were processed with sound analyzing software called *SUGI Speech Analyzer* (Sugitō, 2000). When analyzing the speech extracts, they were divided into tone groups as shown in Table 1. The items measured and identified were: tone at a nucleus, the duration of a tone group and that of a

pause; number of syllables within a tone group, rate of speech (syllables/second), prominent item(s) in a tone group, the item with the highest pitch within a tone group, and the item(s) with maximum sound pressure level within a tone group.

3. Results and Remarks

The results of the measurement were shown in Table 1 and Table 2. in the appendix at the end of this paper. The items of the leftmost column in Table 1 are the first two words of the tone groups. The results of the items identified and measured will be discussed here.

3.1 division of tone groups and tone of nucleus

Watanabe (1994) listed a fragment of speech done by N. Kinnock, the leader of the British Labour Party, in 1992 as a successful speech and one of the typical styles for public speech. *Watanabe* reported that Kinnock divided his speech into as many tone groups as possible and so they became shorter ones without any chaff and with only carefully selected words. Twenty-nine out of thirty tone groups ended with falling tone, many of which were very high falling, with decreasing voice quantity within a tone group for a modulation for effect. As a result Kinnock's speech sounded rhythmical and reassuring.

The material in the present paper was divided into thirty-two tone groups. Obama used nine non-falling tones at the end of the tone groups. They consisted of eight levels and one rising. Since the repetition of (high) falling tone is said to have a reassuring effect, the audience could not have helped concentrating themselves on what the speaker would say thereafter with non-falling tones. In that sense Obama's speech would have sounded attractive and the attraction might have yielded an impression of his eloquence in the mind of the audience.

Nuclear syllables were identified by the author and listed as prominent item(s) in the middle column of Table 1. As for the tones of the thirty-four nuclear syllables, there were twenty fallings, six risings (crisis, violence, many, m measurable, easily, know), five rise-falls (network, hatred, weakened, nation, costly), two levels (ways, America) and one unidentified. According to *Davy (1968)*, which reported the rates of tones at nuclei in spoken British English, the tones consisted of 59% falling, 16% rising, 8% level and 17% others. Although it is difficult to establish a standard ratio of tones for public speeches with limited data to compare with, Obama's usage of tone for nucleus was not so partial as Kinnock's plethora of falling tone for nuclei. Thus, ideal public speech seems to have some styles in the use of tone groups, their ultimate tones, and tones for nuclei.

3.2 duration of utterance (pause and speech)

The total duration of the material was 77,991 msec. President Obama used no fillers such as

“ah-” or “erm-” in the extract and all pauses he used became silent pauses. Table 2 shows the duration of pauses inserted. The total duration of pauses reached up to 21,705 msec., which means President Obama’s speech consisted of 72.2% utterance and 27.8% pause

The preceding studies on silent pause, *Goldman-Eisler (1961)* and *Crystal & Davy (1975)*, reported that most silent pauses used in conversation, discussion or speech were within two seconds and there were few which lasted for much longer than two seconds. Apart from the two pauses lasting a little longer than two seconds, there was one significantly longer than two seconds, which was 3,165msec. pause before the last paragraph. The material consisted of three paragraphs and it would be natural for a speaker to put a comparatively long pause between the paragraphs. The value of the longest pause, 3,165, however, was outstanding and seems to be worth noticing. The last paragraph was the conclusion of the material. To emphasize the last paragraph, Obama is supposed to have used a much longest pause to charge energy enough to finish the last paragraph. It seems to be just like an athlete crouching down before having a big jump or a boxer drawing his dominant arm just before giving the finishing blow.

Mori, Higgins, and Kiritani (2005) examined the frequency of silent pause using the London Lund Corpus of Spoken English (which included a wide variety of monologues and dialogues such as TV and radio broadcasts, political and legal statements, telephone and face-to-face conversations) as materials and reported that a silent pause appeared after every 6.6 words on the average. If 190²⁾, the total number of words of the material of this paper, is divided by 6.6, the average number of words which follows after a pause, it is 28.8, from which we can see that, if President Obama used pauses with the average frequency, the extract from his speech would contain 27.8 (28.8-1) pauses. The number of pauses actually used by Obama was 29 which was close to the averaged value in various types of speech. Apart from his value, however, it should be noticed that there was a great difference in duration between the longest tone group and the shortest tone group and in the number of syllables between the tone group with the most syllables and the tone groups with the fewest syllables. He uttered for nearly 4.0 sec. at the longest and for less than 0.9 sec. at the shortest with one breath. The breath group with the most syllables was “*and each day brings further evidence that the ways we use energy strengthen*” consisting of eighteen syllables and the breath groups with the fewest³⁾ were “*for a new age*”, “*Homes have been lost*”, “*across our land*”, and “*they will be met*” consisting of four syllables each. Comparing the former with the averaged value⁴⁾ of the latter group, the former was uttered with a pace of 1.2 times as fast as the latter in the rate of speech although the former had 4.5 times as many syllables as the latter had. According to the feature of English rhythm, the rate of speech would be increased as a tone group has more syllables. The value 1.2, therefore, seems to be remarkable.

3.3 rate of speech

There are not a few preceding studies on the rate of speech. Focusing on U. S. presidents' speeches, *Heffner (1950)* reported the values of two one-minute samples including pauses of President Franklin D. Roosevelt and those of President Harry S. Truman. The averaged values of the former were 97.5 words/min, 2.3 syllables/sec, and 6.2 phonemes/sec and those of the latter were 155 words/min, 4.0 syllables/sec, and 10.3 phonemes/sec.

Taniguchi (1988) measured some extracted passages including pauses of inaugural addresses of three U. S. presidents, John F. Kennedy, James E. Carter, and Ronald W. Reagan. The averaged values⁵⁾ of Kennedy were 109.5 words/min, 2.8 syllables/sec, and 6.8 phonemes/sec, the values⁶⁾ of Carter were 105.6 words/min, 2.4 syllables/sec, and 6.1 phonemes/sec, and the values of Reagan⁷⁾ were 168.7 words/min, 3.6 syllables/sec, and 9.5 phonemes/sec. Taniguchi adopted a speech in a debate, a few public speeches, a few speeches in TV commercials, and some speeches in movies besides the presidents' inaugural addresses mentioned above, and at the same time carried out a perceptual experiment to know how fast such speeches were perceived by listeners. In the conclusion, it was reported that public addresses were mostly spoken at a slower rate than the normal conversational speed, ranging from 90 words/min to 170 words/min, or from 2.3 syllables/sec to 4.2 syllables/sec, or 5.2 phonemes/sec to 9.7 phonemes/sec, while the acceptable range of normal conversational speech was from 110 words/min to 230 words/min, with a median of 180 words/min, or from 2.7 syllables/sec to 5.8 syllables/sec, median 4.5 syllables/sec, or from 6.0 phonemes/sec to 12.8 phonemes/sec, median 10.0 phonemes/sec⁸⁾. Also it was reported as a part of the conclusion that President Reagan spoke at the perfectly normal conversational rate from a viewpoint of perception. That could be one of the reasons why he was said to be a great speaker.

The results of the present paper were obtained with pauses as processed in the previous section 3.1. The total duration of the material was 77,991 msec. with 190 words, 259 syllables, and 674 phonemes altogether. The rates of Obama's speech were 146.1 words/min, or 3.32 syllables/sec, or 8.64 phonemes/sec. Obama spoke faster than Roosevelt, Carter, or Kennedy and slower than Truman or Reagan. Considering the range of public addresses reported in *Taniguchi*, President Obama seems to have used the ideally normal speed for them.

Looking at the column of syllables/sec in Table 1, most of the values ranged from 4.0 to 6.0. The value at '*is inevitable that*' was 8.12 which was much the largest, but is not worth discussing here as the word '*inevitable*' has as many as five syllables. In the material there was an enumeration of simple-structured sentences expressing what serious problems the United States had: '*Homes have been lost; jobs shed; businesses shuttered. Our health care is too costly; our schools fail too many;*'. The values at the enumeration were comparatively smaller than the other tone groups. And it would be the most remarkable that the second-lowest

value, 2.88, appeared immediately after a large value, 5.20, in the preceding tone group. President Obama seems to slow down at a tone group to emphasize and the large value at the preceding tone group seems to be the cue to start slowing down. The last tone group of the material actually sounded so impressive and the response of the audience reached the peak at this moment. In *Ichizaki (2011)*, however, it was concluded based on the contrasting result that the rate of speech reduced at the tone group preceding the tone group to emphasize, or the sound was lengthened at the item just before the emphasized tone group. So it would be better to say that President Obama is good at modulating the rate of speech for emphasizing.

3.4 prominence, pitch, and sound pressure

In the right-hand side of the prominent items in Table 1, the items having the highest pitch, the items having the greatest pitch change, and the items having the maximum sound pressure level were also listed to examine the relationship between nucleus and such phonetic features. Concerning sound pressure level shown in *SUGI Speech Analyzer*, as it is a unit used for convenience when referring to intensity, it will be described as sound pressure hereafter.

There were some places where fundamental frequency (F_0 henceforth) was not extracted enough to be examined because the vocal cords are apt to vibrate aperiodically especially at the end of a breath group as indicated in *Maekawa (1996)*. Therefore, some cells in the column titled “highest pitch” or “greatest pitch change” in Table 1 were blank (indicated by ϕ).

Thirty-four items having nucleus were recognized for thirty-two tone groups and twenty-seven of the items were situated at the end of tone groups. Excluding four tone groups which had a blank cell in Table 1, twenty-two prominent items out of the twenty-three at the end of tone groups were substantiated as the items having nucleus by either the greatest value of pitch, pitch change, or sound pressure: five items were supported by all three values, five items by pitch change and sound pressure, three items by the value of pitch and pitch change, eight items by pitch change only, and one item by sound pressure only. Only one item, “*hatred*,” out of twenty-three items having nucleus which appeared at the end of tone group was supported by none of the three features. The tone group having “*hatred*” had another prominent item having nucleus, “*violence*.” The highest pitch of “*hatred*” was only 6 Hz lower than that of “*violence*” and the maximum sound pressure of the former was also only 1 dB smaller than that of the latter. These results clearly show that prominence is made from high pitch, wide pitch change, great sound pressure, or long duration.

Apart from such phonetic features, it was remarkable that the first tone group of the last paragraph, which was the concluding paragraph of the material, showed the maximum

sound pressure throughout the material at each of four content words, “*Today, say, challenges, real*”. Listening to the sentence having maximum sound pressure after the longest pause throughout the material, the audience must have been greatly attracted by the speaker at that moment.

4. Conclusion

Non-falling tones were used at the end of nearly a third of the tone groups and there was a preference for non-falling tones with a rate of nearly 40% at nuclear syllables. The use of non-falling tones seems to be somewhat excessive for public addresses. Speech occupied approximately 72% while pause occupied approximately 28%. Concerning the frequency and the duration of pause, President Obama's speech was just like people's daily conversation, except for a slightly slower rate than normal conversation (which was a standard value for public speeches). Also, few fillers were used throughout the whole speech lasting nearly twenty minutes, which may have been partially supported with the aid of a teleprompter set on either side of him. His division of tone groups was also neither more nor less than that in ordinary conversation. As a result his utterance without faltering or stammering seems to have sounded fluent and natural.

On the other hand there were some particular features of President Obama. The tone groups he used varied widely both in duration, 862msec. – 3,975msec. and in number of syllables, 4 – 18, although there was not much difference in the rate of speech. He appropriately put a lot of information or a little information into a tone group. It could be said that he is good at modulating the quantity of information within a tone group according to the circumstances. Another feature is the rate reduction at the tone group to emphasize and the rate elevation at the tone group preceding it. Besides, it must be another feature for him to use maximum sound pressure successively in the first tone group of the concluding paragraph. Skillful usages of such phonetic features seem to be President Obama at his best.

The target material of this paper was just a fragment of an inaugural address made by President Obama. The writer's next research is being planned to be reinforced with more materials and it is hoped to show more decisive results.

Notes:

- 1) The speech material was extracted from
<http://www.youtube.com/watch?v=DEGd75awApM>
- 2) “*far-reaching*” in the material was recognized as two words.
- 3) The tone group with the fewest syllables, “*jobs shed,*” had no pause thereafter and was

linked to the following tone group, so it was not recognized as a breath group.

- 4) It was 3.725 syllables/sec.
- 5) As *Taniguchi* adopted three extracts, whose durations were 24.54sec., 18.26sec., and 7.52sec., as the material to measure, the averaged values listed here were calculated considering the ratio of duration.
- 6) The duration of the extract was 28.99 sec.
- 7) The duration of the extract was 19.56 sec.
- 8) *Taniguchi* showed rate of speech using syllables/min and phonemes/min and they were shifted to syllables/sec and phonemes/sec.

References

- [1] D. Crystal and D. Davy. *Advanced Conversational English*. London: Longman, 1975.
- [2] D. Davy. "A study of intonation and analogous features as exponents of stylistic variation, with special reference to a comparison of conversation with written English read aloud," University of London M. A. thesis, 1968.
- [3] F. Goldman-Eisler. "The rate of changes in the rate of articulation," *Language and Speech*, no. 4, pp. 171-174, 1961
- [4] R-M. S. Heffner. *General Phonetics*. University. of Wisconsin Press, 1950.
- [5] K. Ichizaki. "A pilot study of the eloquence of President Obama from an acoustic point of view: 'World without nuclear weapons' speech in Prague, Czech Republic," *The Bulletin of Miyazaki Gakuen Jr. College*, vol. 3, pp. 1-11, 2011.
- [6] K. Maekawa. "Onnkyoonnseigaku no tennbo," *The Bulletin of the Phonetic Society of Japan*, no. 211, pp.12-19, 1996.
- [7] Y. Mori, A. Higgins, and S. Kiritani. "Functions and locations of long silent pauses in English discourse," *JACET Bulletin* , no. 41, pp.91-110, 2005.
- [8] M. Sugito. *SUGI Speech Analyzer CD-ROM*, Yokohama: Animo, 2000.
- [9] M. Taniguchi. "Average range and limitations of English speech velocity," *The Bulletin of Chikushi Jogakuen University Jr. College*, vol. 23, pp. 55-82, 1988.
- [10] K. Watanabe. *Eigo intonation ron*. Kenkyusha, 1994.

Table 1 : Phonetic Features Analyzed on Extract

sylls = no. of syllables, R. of Sp. = rate of speech (syllables/sec),

H. P. = item with highest pitch, G. P. Ch. = item with the greatest pitch change,

M. P. = item with the maximum pressure

tone group	duration (ms)	sylls	R.of Sp.	prominent item	H. P.	G. P. Ch.	M. P.
<i>That we...</i>	1,860	9	4.84	crisis	ϕ	ϕ	we, midst, crisis
<i>is now...</i>	1,212	6	4.95	understood	now	ϕ	well
<i>Our nation...</i>	1,190	6	5.04	war	nation	war	war
<i>against a...</i>	1,633	8	4.90	network	far	network	far
<i>of violence...</i>	1,160	7	6.03	violence, hatred	violence	violence	violence
<i>Our economy...</i>	3,104	14	4.51	weekend	weekend	weekend	consequence
<i>greed and...</i>	2,132	14	6.57	some	greed	some	greed, irresponsibility
<i>but also...</i>	3,029	14	4.62	choices	choices	choices	also, collective
<i>and prepare...</i>	1,059	6	5.67	nation	prepare	nation	prepare
<i>for a...</i>	911	4	4.39	age	new	age	new
<i>Homes have...</i>	1,221	4	3.28	lost	Home	lost	lost
<i>jobs shed;</i>	884	2	2.26	shed	shed	shed	shed
<i>businesses shuttered.</i>	972	6	6.17	shuttered	shuttered	shuttered	shuttered
<i>Our health...</i>	1,808	7	3.87	costly	costly	health	care, costly
<i>our schools...</i>	1,589	5	3.15	many	many	many	school, fail
<i>and each...</i>	3,975	18	4.53	ways	each	ways	day, evidence, ways, energy
<i>our adversaries...</i>	2,154	10	4.64	planet	adversaries	planet	our
<i>These are...</i>	2,027	10	4.93	crisis	These	ϕ	are, the
<i>subject to...</i>	2,009	9	4.48	statistics	statistics	statistics	statistics
<i>Less measurable;</i>	1,007	5	4.97	measurable	measurable	measurable	Less, measurable
<i>but no...</i>	3,104	13	4.19	confidence	no, sapping	ϕ	sapping
<i>across our...</i>	919	4	4.35	land	our	land	across, our
<i>a nagging...</i>	1,541	5	3.24	fear	fear	nagging	fear
<i>America's decline</i>	1,230	6	4.88	decline	decline	decline	America's
<i>is inevitable...</i>	862	7	8.12	inevitable	inevitable	inevitable	inevitable
<i>the next...</i>	2,237	11	4.92	sights	generation	sights	lower, sights
<i>Today I...</i>	3,126	15	4.80	real	Today	real	Today, say, challenges, real
<i>They are...</i>	2,031	9	4.43	many	serious	many	serious
<i>They will...</i>	1,917	9	4.69	easily	easily	easily	not, easily
<i>or in...</i>	1,450	6	4.14	time	short	time	time
<i>But know...</i>	1,345	7	5.20	know, America	know	know	But, this, America
<i>they will...</i>	1,387	4	2.88	will	will	will	will
Total ; Average	56,286	259	4.60				

Table 2 : Duration of Pause (msec.)

1 st pause (no breather)	179
2 nd pause	1,208
3 rd pause	442
4 th pause	455
5 th pause	1,225
6 th pause	332
7 th pause	884
8 th pause	613
9 th pause (no breather)	201
10 th pause	1,317
11 th pause (no breather)	240
12 th pause	534
13 th pause	547
14 th pause	253
15 th pause	2,267
16 th pause	643
17 th pause	2,219
18 th pause	415
19 th pause	630
20 th pause	840
21 st pause (no breather)	385
22 nd pause	424
23 rd pause	696
24 th pause	3,165
25 th pause	503
26 th pause	214
27 th pause	389
28 th pause	258
29 th pause (no breather)	227
total	21,705