

A Computer Tool in the Study of Tones¹

杜 彩 春

(作者為本校西語系校友)

摘 要

這篇文章介紹了一種應用電腦及統計方法分析語言聲調的工具，電腦方面是利用硬體的配合，研究者可將聲音輸入電腦，軟體程式將聲音的基本頻率 (Fundamental Frequency) 算出並顯現在電腦顯示器上，統計方法指的是一種叫 K-means 的互除法，這個互除法可將一大堆資料根據某種變數分門歸類，每個類別裏的資料都是具有相似的某種成份。

本文以臺灣話為實驗對象，一共有三個實驗，第一個實驗：先用電腦分析一二八個單音節字的聲調，然後把每個音節聲調的起點、中點、終點頻率，起點至終點，起點至中點，中點至終點的斜率及調長等七個參數算出並存在一個檔案，接著根據這七個參數的結合將一二八個音節以 K-means 分為七類 (根據過去的研究臺灣話有七個聲調)，分類的結果，我發現當一個音節單獨念時一共有六個聲調：五個非入聲調及一個入聲調。第二個實驗分析變調，作者以電腦分析一二八個雙音節詞語，然後將第一音節的七種參數算出存在檔案，接著以 K-means 方法分類所有第一音節的聲調，結果顯示一共有七個變調，入聲音節有二個不同變調。

第三個實驗是調查在 /a/「仔」字前的聲調，利用前二個實驗方法分類「仔」字前調類，作者發現若「仔」字前的音節乃指人名或親屬稱謂，這個音節不變調，此外由此音節調值，作者找出了入聲字的本調，一是高平調另一是低降調。

作者以電腦及統計程序為工具研究聲調現象。發現臺灣話裏在深層結構中有七個聲調，五個非入聲調及二個入聲調，在表層結構這二個入聲調合而為一，除此之外這項工具或多或少可幫助一些難以聽覺判辨聲調者研究語音現象。

ABSTRACT

This paper introduces a method which incorporates statistics and computer techniques in the investigation of the inventory of Taiwanese tones. The result turns out to be fairly satisfactory.

0. Introduction

With the development of digital computer techniques, people doing acoustic analysis such as Tseng (1981) have been able to use computers to analyze speech sounds, abstracting suprasegmental features like stress, intonation, and tone. Most researchers, in studying the tonal phenomena, derive the tone inventory of a tone language through perception or sonograph machine. Perception is, of course, more subjective than objective. I do not claim that it is incorrect to count on perception in the study of tones, particularly since prior to the development of computer technique or sonograph machine, it was the only method available. However, it is not easy for tone-deaf people to perceive contrastive tones. Sonograph machines have already been available for waveform and frequency display. However, in the calculation of the fundamental frequency, it is not very convenient, e.g. the F_0 is usually obtained by dividing the 10th harmonic frequency by 10. With the help of the computer hardware and software, the F_0 can be calculated automatically by computer program. In this paper, I will introduce a tool that can be used to derive the tone inventory of a language without depending on subjective perception or measuring frequencies through sonograph machine.

The tone-language I used for the experiment is a Chinese language, Taiwanese which is spoken in Taiwan.² Basically, each syllable in Taiwanese has two tones: the citation form which occurs in isolation or before a neutral tone and the sandhi form which occurs in other environments³. Except for the Qingshue dialect mentioned briefly by Ho (1984) which has six citation tones, all studies in Taiwanese tones (R. Cheng 1973, 1977, Tung et al 1967, Zhang 1983) report that there are seven citation tones. Of the seven tones, five are non-entering tones and two are entering tones. The non-entering tones are with open syllables or syllables ending in nasal consonants, e.g./si55/'silk', /si33/'yes', /si53/'die', /si21/'four', /si13/'time' in R. Cheng (1977). The entering tones are with syllables ending in stops /p, t, k, ?/, e.g./sit53/'honest', /sit21/'lose' in R. Cheng (1977). The former are usually longer than the latter. In the Qingshue dialect, there are five non-entering tones but only one entering tone. I am a native speaker of Taiwanese. In this paper, I confirm Ho's findings that the Taiwanese spoken in some areas in Taiwan, like Qingshue dialect, does have five non-entering citation tones and only one entering citation tone. In order to find out whether there are actually six citation tones, I used a computer and a pattern-matching algorithm to derive the inventory of the citation tones and the sandhi tones. From the sandhi tones, it becomes clear there are six citation tones on the surface. The two entering tones are neutralized

in isolation but the distinction remains in the underlying forms. In the following sections, I will first describe the motivation for the experiment. Then I discuss the tool used in the analysis of the tones: the computer hardware and software used to extract the fundamental frequency of each syllable and the statistical algorithm used to classify the tones. Finally, I will present the statistical results from which the tone inventory is derived.

1. Motivation

The idea of using computers and statistical algorithm to study tones came from research in Mandarin tones done by C. C. Cheng (1982) who extracted six parameters from each syllable input through a pitch extraction device interfaced with PLATO. These six parameters are the beginning frequency (BF), the mid frequency (MF), the end frequency (EF), the beginning-end slope (BESL), the beginning-mid (BMSL) slope, and the mid-end slope (MESL). The six parameters are used for tone recognition. The recognition result shows that some parameters or combination of parameters are better than others in identifying tones. With the assumption that some parameters are better cues in tone recognition and can be used to group syllables produced with one tone within a person's pitch range together, I adopted these six parameters for the use of tone classification. Furthermore, I added a seventh parameter, tone duration (DUR) because tone length between the non-entering tones and entering tones is distinctive. The classification of the citation tones is based on a pattern matching process called the K-means algorithm (Macqueen 1965). This process groups syllables that have similar parameter values together. The average parameter values of each group represent one tone value. To obtain the values of the seven parameters, I used a speech program to get the fundamental frequency of each syllable. The seven parameter values were calculated and stored on a file for the subsequent grouping process.

2. Computer tools

2.1 The speech program

In examining the tones, I used a digital computer technique to calculate the fundamental frequency of the speech input. In this section, I will briefly discuss the structure of this speech program.

The computer program used for the speech processing of voice in the pitch

extraction experiment requires an electronic analog-digital (A/D) and digital-analog (D/A) converter.⁴ The program was written in BASIC and was designed to do the following:

- (a) accept speech input, display the waveform, and echo the sound back to the user;
- (b) splice the speech input in order to extract monosyllabic words;
- (c) calculate and plot the fundamental frequency of the speech input.

With this A/D and D/A converter, the program sampled at the rate of 10K/sec for male voices and 20K/sec for female voices.⁵ Using this program, sounds could be input through a tape recorder, the waveform displayed on the computer screen, and the sound reproduced. The sound was 1.2 seconds long for male voices and 0.62 seconds long for female voices.

2.2 Extraction of the fundamental frequencies

The fundamental frequency estimation (also called pitch period estimation) was carried out by using the autocorrelation function (Rabiner and Schafer 1976: 141-148), the center clipping function (Sondhi 1968), and the three-level center clipping function (Dubnowski, Schafer, and Rabiner 1976).⁶ With these functions, the software was written to calculate the Fo of a monosyllabic input sound and print the frequency on the computer screen. Figure 1 shows the waveform and fundamental frequency of syllable /si55/ 'silk' with high level tone computed by the speech program.

The seven parameters were obtained in the following way. The program was also written to search for the meaningful beginning and end fundamental frequency points of a syllable. In the process of searching for the meaningful beginning point, the difference in the fundamental frequency between two neighboring points is checked until a difference in frequency not greater than 10Hz is found. Then the leftmost of these two points was selected as the beginning point. Next the program searched for the meaningful end point from the end of the syllable. The search continued backwards (leftwards) until a difference in frequency not greater than 10Hz was found. Then the right point was chosen as the end point. Once the beginning and end points were located, the time of each point was calculated. The tone length was obtained by subtracting the time of the beginning point frequency from the time of the end point frequency. The mid point time and frequency were determined as well. The parameter 'duration' in our experiment refers to tone duration rather than syllable duration. To get the beginning-end slope, the

A Computer Tool in the Study of Tones

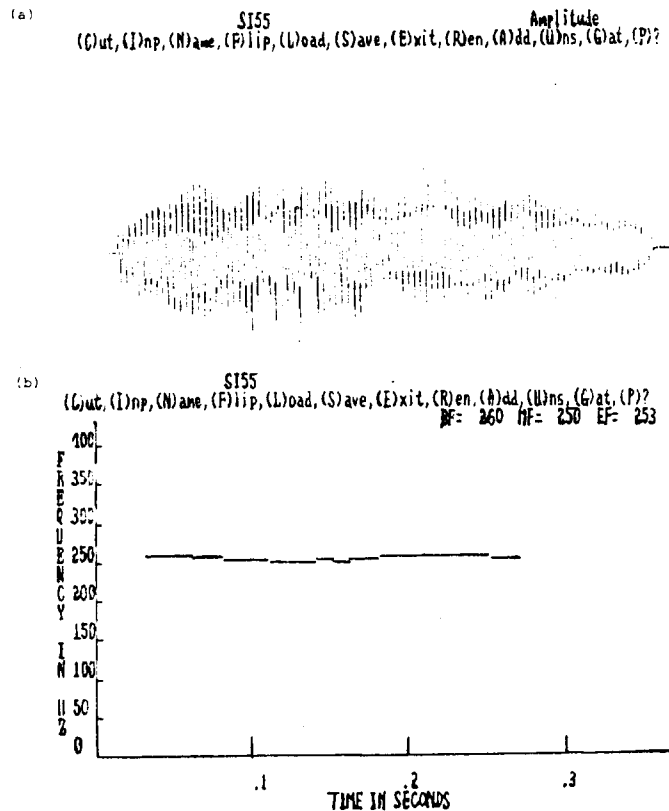


Figure 1. Waveform and fundamental frequency of syllable /si/ 'silk' in high level tone.

beginning-mid slope, and the mid-end slope, C. C. Cheng's formulae (1979) were followed:⁷

$$\begin{aligned} \text{beginning-end slope} &= (\text{end frequency} - \text{beginning frequency}) / (\text{end time} - \text{beginning time}) \\ \text{beginning-mid slope} &= (\text{mid frequency} - \text{beginning frequency}) / (\text{mid time} - \text{beginning time}) \\ \text{mid-end slope} &= (\text{end frequency} - \text{mid frequency}) / (\text{end time} - \text{mid time}) \end{aligned}$$

The seven parameters of each syllable were stored on a file for the purpose of similarity grouping discussed next.

2.3 K-means algorithm

The K-means algorithm is a process of partitioning large groups of samples

into k sets on the basis of specified values. MacQueen (1965:283) gives a good definition of the K-means process. He says.

Let Z_1, Z_2, \dots be a random sequence of points (vectors) \dots . The K-means procedure consists of simply starting with K groups, each of which consists of a single random point, and thereafter adding each new point to the group whose mean the new point is nearest. After a point is added to a group, the mean of that group is adjusted in order to take account of the new point. Thus at each stage the K-means are, in fact, the means of the groups they represent.

Basically what the K-means algorithm does is to cluster a body of samples into k number of groups, each of which consists of samples of similar X . X can be anything, e.g., social values, distance, etc. The steps for clustering N samples are the following:

- (a) select the number of groups, i.e. K ;
- (b) select (randomly) a sample as the initial center of each group; whatever value the initially selected sample has is considered initial mean value of the group;
- (c) compute the difference between the value of any new sample and that of each group center, then add this new sample to the group to whose mean the new sample is nearest;
- (d) recalculate the mean and go back to step c until all the samples are properly classified.

There are several applications of the K-means process; one of them is 'similarity grouping' (MacQueen, 1965:288). In my experiment, this K-means process is used to do similarity grouping, i.e. classify all the speech data into seven tonal groups based on the distances among the parameters.

3. Experiment

As mentioned, previous studies in Taiwanese tones (R. Cheng 1977, Tung et al 1967, Zhang 1983) show that there are seven citation tones. Thus in the very beginning I will also assume that there are seven citation tones in the data to be examined. This section is organized as follows. Section 3.1 deals with these citation tones. Section 3.2 and 3.3 deal with sandhi tones.

3.1 Citation tones

3.1.1 Data

The speech data were citation forms, consisting of 140 monosyllabic words selected from R. Cheng (1977) and Zhang (1983), 20 words for each tone. In order to differentiate these words, each syllable was transcribed with a numerical tone value. For example, /Si55/ was a syllables with high level tone. Here I adopt R. Cheng's tone system to transcribe these data. In R. Cheng (1977), the seven lexical tones are five nonentering tones 55, 53, 33, 21, 13 and the two entering tones 53 and 21. In order not to mix the two non-entering tones 21 and 53 with the corresponding entering tones 21 and 53, in the clustering program later, I used 43 instead of 53 for the non-entering tone 53 and 22 instead of 21 for the entering tone 21. In my experiment, for the non-entering (long) tones, 55 is used for high level tone, 33 for mid level tone, 43 for high falling tone, 21 for low falling tone, and 13 for rising tone.⁸ For the entering (short) tones, 53 is used for high falling tone and 22 for low falling tone.

Two speakers read the data. They are both bilingual, native speakers of Taiwanese in their twenties. Speaker A is from the Kaoxiong area and speaker B, the author, is from the Taichung area. The two speakers recorded the data three times during a three-week period. They read the word list at a normal speed. The speech was recorded on a reel tape using a REVOX PR99 recording system. The recorded speech then was transferred from reel to cassette using a SANYO RD5030 stereo cassette recorder. Using the speech program described in section 2.1, I obtained the fundamental frequency of each monosyllabic word in the recording. Seven parameters were extracted from each word and were stored on a file for the purpose of similarity grouping. Since both speakers read the data three times, the final file used for the grouping test for each speaker was the average of the parameters obtained for each of the speech words in the three recordings.

3.1.2 Similarity grouping

All the parameters of the 280 words read by the two speakers were normalized at first within the range of 0 to 100. Thus, e.g. the highest frequency of speaker A's data (275.67Hz) and the highest frequency of speaker B's data (248.50 Hz) were both set equal to 100. The lowest frequency of speaker A's data (149.00 Hz) and the lowest frequency of speaker B's data (133.67 Hz) were both set equal to

0. The remainder of the frequencies were normalized within this range.⁹ Since we assumed that there were seven citation tones, the program was written to cluster the data into seven groups. These data were clustered on the basis of the closeness of the parameters selected for similarity grouping. The K-means program first selected one word from each tone to be the initial center of each group. The mean values of each group depend on which parameters are selected as the clustering parameters. For example, if parameters "beginning point frequency" (BF) and "end point frequency" (EF) are the clustering parameters, then the BF and the EF of the initially selected group center are the mean values of that group. Next the BF and EF parameters of a second word were compared with those of the seven groups. This second word was added to the group whose mean values that the BF and EF of this word were nearest to, i.e. to the group which had the smallest distance between the parameters of the second word and the mean values of a group. The distance was calculated by the pythagorean theorem: the distance between two points equals to the square root of the sum of the length of the sides. For instance, if I assume that the BF's and EF's of the center of one of the groups and the second word are (x_1, y_1) and (x_2, y_2) respectively, then the distance between the group and the second word is the square root of $(x_1 - x_2)^2 + (y_1 - y_2)^2$. Next a third word was selected and added to a group based on the smallest distance difference between the parameter values and the group mean values. After all the 280 speech words were classified, new means were obtained for each group by first adding the clustering parameter value of each word collected in a group and then dividing the sum over the number of words in this group. For example, if BF and EF are the clustering parameters, then the BF values of all the words in the group are added together, then divided by the number of the total words in the group. Similarly the EF values of all the words are added, and the result is divided by the total number of the words in the group. These two newly averaged BF and EF values become the new center mean values. With these new means, all the data were classified again. This recalculation of group mean values and reclassification was repeated until the last recalculated mean values were the same as the previous ones. This means that the words clustered together in each group are stable. Words clustered in the same group can be said to have similar parameter values. The program then printed out the number of words and the words in each group. Since the grouping can be based on just one parameter, two parameters, or up to seven parameters, there are in total 127 combinations of the seven parameters. From this print-out, it is generally easy to tell which parameter or combinations of parameters can best distinguish among these seven

tones. In the next section, we will discuss what the statistics can tell us about the citation tones in Taiwanese.

3.1.3 Result and discussion

In total, there are 127 classification results for the 280 experimental syllables based on the combination of the seven parameters. Among these 127 results, we picked one classification shown in Table 1 based on four parameters: the beginning frequency, the mid frequency, the end frequency, and the tone duration for discussion. The combination of these four parameters identified the citation tones best.¹⁰

In Table 1, we observe the following:

The citation forms of 39 syllables transcribed with 55 are collected in group 1. The citation forms of 40 syllables transcribed with 43 are collected in group 2. The citation forms of 40 syllables transcribed with 33 are collected in group 3. The citation forms of 35 syllables transcribed with tone 22 and 35 syllables transcribed with tone 53 are collected in group 4. The citation forms of 18 syllables transcribed with 21 are collected in group 5 and the remaining 22 syllables are collected in group 6. The citation forms of 40 syllables transcribed with tone 13 are collected in group 7.

The first three center mean values in each group in Table 1 are the normalized average frequency values for the beginning, the mid, and the end points. I convert each of these three mean values into the five-point scale tone system (Chao 1930). The normalized highest mean value 86.054 is interpreted as 5 and the lowest normalized mean value 11.351 is interpreted as 1. The pitch range from 5 to 1 is derived by first dividing the difference between 86.054 and 11.351 by 5, making an interval of 14.941. Next we subtract 14.941 from 86.054, which is the range for 5. Again we subtract 14.941 from 71.112, which is the range for 4, etc. The mean value range for each point in the scale is summarized in Table 2.

Table 1: citation tone grouping using parameters BF, MF, EF, DUR

group no. 1 contains 39 members
84.971 78.991 86.054 70.333
55:39 33:0 43:0 21:0 13:0 22:0 53:0

group no. 2 contains 40 members
84.632 60.900 26.712 37.480
55:0 33:0 43:40 21:0 13:0 22:0 53:0

group no. 3 contains 43 members

65.682 56.900 63.771 73.799

55:1 33:40 43:0 21:0 13:0 22:0 53:2

group no. 4 contains 70 members

61.350 50.626 33.835 12.928

55:0 33:0 43:0 21:0 13:0 22:35 53:35

group no. 5 contains 26 members

53.457 36.214 18.419 17.281

55:0 33:0 43:0 21:18 13:0 22:5 53:3

group no. 6 contains 22 members

50.476 26.290 11.351 48.411

55:0 33:0 43:0 21:22 13:0 22:0 53:0

group no. 7 contains 40 members

37.414 22.707 48.996 78.042

55:0 33:0 43:0 21:0 13:40 22:0 53:0

*The decimal numbers beneath each group are the group center mean values.

Table 2 Tone letter assigned to each group in Table 1

numerical system		center mean values	
5	from	86.054	group 1 55
	to	71.113	group 2 52
4	from	71.112	group 3 44
	to	56.172	group 4 42
3	from	56.171	group 5 31
	to	41.231	group 6 31
2	from	41.230	group 7 213
	to	26.291	
1	from	26.290	
	to	11.351	

Using Table 2, I can see that the experimental clustering groups represent six tone values: 55, 52, 44, 42, 31, and 213. Syllables transcribed with 21 distributed in two groups 5 and 6, but they were all produced with tone 31. The reason that the 18 syllables transcribed with 21 in group 5 were differentiated from the remaining 22 syllables transcribed with 21 in group 6 is that group 5 syllables were produced with shorter tone durations. However, this discrepancy will not influence the statistical result. Group 4 contains most of syllables in the entering

tones, which suggests that phonetically, there is only one entering tone. Later, in examining tone sandhi, I will show that underlyingly there are two entering tones. Another noteworthy item is that the duration mean values tell us that the entering tones are comparatively shorter than non-entering tones.

It is clear that the Taiwanese dialect R. Cheng (1977) studies has different tone values from my experimental data. The correspondence between the tone letters in the transcription and the derived tones from Table 2 is listed in Table 3.

Table 3

transcription		derived
55	↔	55
33	↔	44
43	↔	52
21	↔	31
13	↔	213
53	↔	42
22	↔	42

* ↔ means 'corresponds to'

3.2 Sandhi tones

3.2.1 Data

We made up 140 two-syllable phrases, each of which consisted of a monosyllabic word used in the previous analysis followed by a second monosyllabic word. There were 20 phrases for each tone. As before, all the monosyllabic words in the phrases were transcribed with the citation tone letters used in the previous experiment. The first syllable of each phrase was analyzed by the speech program. The seven parameters were extracted and stored. All the parameters were normalized between 100 and 0. Then, based on the combination of these 7 parameters, these 280 tokens were classified by the K-means program in the same way as described in section 3.1.2.

3.2.2. Result and discussion

One hundred and twenty-seven classifications were obtained. It was found

that parameters related to mid point frequency or slope do not play an important role in the two-word phrase classification. The best classification, then, is shown in Table 4, based on BF, EF, and DUR.

Again, the normalized mean beginning and end frequency values for BF and EF of all seven groups were converted into the numerical equivalents. These converted tone values from Table 4 are listed in Table 5.

Table 4 sandhi tone grouping using parameters BF, EF, and DUR

group no. 1 contains 45 members
88.616 35.576 42.987
55:0 33:0 43:0 21:37 13:0 22:8 53:0
group no. 2 contains 34 members
76.080 59.557 7.040
55:0 33:0 43:0 21:1 13:0 22:29 53:4
group no. 3 contains 34 members
71.406 67.386 64.305
55:3 33:0 43:31 21:0 13:0 22:0 53:0
group no. 4 contains 36 members
50.374 46.753 73.917
55:13 33:0 43:9 21:0 13:14 22:0 53:0

Table 5

group 1	52
group 2	54
group 3	44
group 4	33
group 5	32
group 6	21
group 7	21

In Table 4, some syllables transcribed with 22 and 53 have not only long tone durations but different sandhi forms from the rest syllables transcribed with the same tones. I will discuss this later. At this moment, I will ignore syllables transcribed with entering tones which are long. According to Tables 4 and 5, we have the following sandhi rules:

- (1) Syllables transcribed with 55 and 13 had sandhi tones 33. The difference between 33 and 32 is not phonologically important.

A Computer Tool in the Study of Tones

- (2) Syllables transcribed with 43 had sandhi tone 44.
- (3) Syllables transcribed with 21 had sandhi tone 52.
- (4) Syllables transcribed with 33 had sandhi tone 21.
- (5) Syllables transcribed with 53 had sandhi tone 21.
- (6) Syllables transcribed with 22 had sandhi tone 54.

Using the derived citation tones from the last experiment, I rephrase the sandhi rules in (1) through (6) in Table 6.

Table 6

citation tones		sandhi tones
55,213	⇒	33/32
52	⇒	44
44	⇒	21
31	⇒	52
42 (short)	⇒	54 (short)
42 (short)	⇒	21 (short)

* ⇒ means the tone at the left becomes the tone at the right.

There are six sandhi tones. Citation tone 42 has two sandhi forms 54 and 21. Is there only one underlying entering tone which takes two sandhi forms? If the historical development of tone system in Chinese is taken into consideration, underlyingly there are two entering tones. Historically, the voicing split of the syllable-initial consonants in Chinese increased the tone inventory from 4 tones to 8 tones. Each of the middle Chinese 4 tones, Pin, Shang, Qu, and Ru split into Yin and Yang, with Yin having high tone and Yang, low tone. From a diachronic point of view, there should be two entering tones. This distinction is found in the sandhi forms but not in the citation forms in our data. We can postulate that underlyingly there were two entering tones. Later in the language development process, the two entering tones were neutralized in isolation in our data. Thus the sandhi forms serve as a criterion for the determination of the underlying entering tones. Nevertheless, because the distinction is maintained in the sandhi forms, Ho (1984) who studies Qingshue, a Taiwanese dialect that has one citation entering tone and two sandhi entering forms, proposes that the sandhi forms are actually the underlying forms and the citation form is a derived form. (As does

Lu 1986) This postulation simplifies the prediction of sandhi forms. If the citation form is the underlying form, it would be difficult to predict when it is changed into one tone and when, the other. On the other hand, if the sandhi forms are the underlying forms, then the two tones are neutralized in isolation. Nevertheless, I cannot adopt Ho's view to assume that in my data the sandhi forms are the underlying forms. The evidence comes from another experiment of the tone sandhi before the diminutive /a/, which is discussed below.

3.3 Tone sandhi before diminutive /a/

Unlike the tone sandhi discussed in 3.2, a syllable immediately preceding the diminutive /a/ undergoes a different tone sandhi. R. Cheng (1977) notes that the syllable tone before the diminutive /a/ undergoes the sandhi discussed in section 3.2 first, and then any non-level sandhi tone becomes level. Thus if I follow his sandhi prediction, in my data, 55 becomes 33, 44 becomes 22, 52 becomes 44, 31 becomes 55, 213 becomes 33, 42 becomes either 22 or 55. I used the computer tools to analyze the tones of 56 syllables immediately preceding the diminutive /a/. These data were read by the two speakers who read the other data. The data were transcribed with R. Cheng's tone letters because the derived tone 42 cannot differentiate the entering tones.

Unlike the citation and sandhi tone classification, I was unable to find any good grouping results for these syllables. Every group in each classification consists of syllables in different tones. However, after converting the normalized group mean values of the beginning point frequency and the end point frequency into tone letters, I found some tonal patterns for these syllables. The classification result based on "the beginning frequency", "the end frequency", "the beginning-end slope", and "the tone duration" parameters is shown in Table 7. The converted tone value for each group is shown in Table 8.

Table 7: tone grouping of syllables immediately preceding the diminutive /a/ using parameters BF, EF, BESL, DUR

group no. 1 contains 5 members
88.5193 94.5828 86.8766 47.4167
55:3 33:0 43:1 21:1 13:0 22:0 53:0

group no. 2 contains 11 members
71.8163 73.5996 84.7418 31.0417
55:2 33:0 43:2 21:0 13:0 22:5 53:2

A Computer Tool in the Study of Tones

group no. 3 contains 9 members
 68.8515 69.5113 85.8303 65.381
 55:1 33:3 43:1 21:3 13:0 22:1 53:0

group no. 4 contains 11 members
 44.8631 75.1378 95.7274 33.6458
 55:3 33:3 43:0 21:0 13:0 22:0 53:5

group no. 5 contains 8 members
 42.2399 28.3257 77.1650 24.1667
 55:0 33:0 43:0 21:0 13:0 22:6 53:2

group no. 6 contains 7 members
 41.6362 62.5795 92.114 64.6583
 55:1 33:2 43:0 21:0 13:2 22:0 53:2

group no. 7 contains 6 members
 41.7614 82.9315 94.7355 78.1548
 55:0 33:1 43:0 21:0 13:2 22:0 53:3

Table 8

group 1	55
group 2	44
group 3	44
group 4	24
group 5	31
group 6	23
group 7	25

Basically three types of sandhi tones are derived: level, rising, and falling. The sandhi tones that each citation tone in my data is changed into are summarized below.

citation tones

55	→ 55, 44, 24, 23
44	→ 44, 25, 24, 23
52	→ 55, 44
31	→ 55, 44
213	→ 25, 23
42	→ 23, 25, 31, 44
42	→ 44, 31,

Apparently, the derived tones in my data are different from those in R. Cheng (1977). At the first glance, it seems that one citation tone can have at most three sandhi tones. However, when the semantics of the syllable before /a/ is taken into consideration, it becomes clear that if a syllable before /a/ refers to a proper name, e.g. /a 52 ts'iu55 a52/, it does not undergo the tone change.¹¹ In my data, there are 14 syllables transcribed with tones 55, 33, 53, and 22 referring to proper names. Four syllables transcribed with 55 have either 55 or 44 sandhi tones. Two syllables transcribed with 33 have 44 sandhi tone. In speaker A's speech, two syllables transcribed with tone 53 have 44 sandhi tone and two syllables transcribed with tone 22 have 31 sandhi tone, and in speaker B's speech these four syllables (transcribed with tones 53 and 22) have 31 sandhi tone. Regardless of the phonetic variations during speech production, these syllables keep their citation tones except for syllables transcribed with tone 53. As mentioned in section 3.1, phonetically there is only one entering tone. But for syllables immediately preceding the diminutive /a/, speaker A from Kaoxiong produced syllables in one citation tone with two phonetic tones. Apparently, this speaker makes a phonemic distinction between the two entering tones in this morphological structure. The different tone values that the syllables transcribed with 53 and 22 have before the diminutive /a/ suggest that in my data, there are two entering tones underlyingly: one is level and the other is falling. This phonemic distinction is still found in speaker A's but not B's speech. As a result, I do not follow Ho, (1984) and Lu (1986) in assuming that sandhi tones are the underlying tones.

4. Tone Duration

As noticed in Tables 1 and 4, when tone duration was one of the clustering parameters, some syllables transcribed with entering tones both in the citation and sandhi environment were collected with syllables transcribed with non-entering tones. Checking the average values of tone duration, I see that these syllables were collected in groups with large duration values. This indicates that these syllables were not produced with short tones but rather with long tones. In Table 9 I list some of these syllables and their tone durations in real time.

Table 9

citation form			sandhi form		
0?53	'study'	224ms	tsio53	'stone'	149ms
tsio53	'stone'	179ms	ti?22	'drop'	137.65ms
guat53	'moon'	159ms	gik53	'jade'	142.33ms

R. Cheng (1977) mentions that some syllables ending in the glottal stop drop the final consonant and are produced with a long tone in the sandhi environment, though they still keep the same entering sandhi tone values as those syllables that do not drop the final consonants. In our data, not only /ʔ/ but other final consonants are dropped in both isolation and sandhi environments for some syllables. These syllables, after losing the consonant ending, have different tone values. For example, 8 syllables transcribed with tone 53 and 22 in isolation had tone 31 rather than tone 42 and 8 syllables transcribed with tone 22 in the sandhi context had tone 52 rather than 54. Here is a situation where the entering tones are beginning to disappear and redistribute themselves among the non-entering tones. These syllables are just the first few to undergo this change.

5. Conclusion

In this paper, I introduce a method which incorporates computer techniques and a statistical algorithm to perform the classification of tones in Taiwanese. MacQueen (1965:288) says, 'the purpose of K-means is not to find some unique definitive grouping, but rather to simply aid the investigators in obtaining qualitative and quantitative understanding of large amounts of N-dimensional data by providing them with reasonably good similarity groups.' As shown in this paper, the classification results have helped me discover the tone inventory of and possible phonetic tone sandhi in my Taiwanese speech data. These results also suggest a language development process which gives me an idea of how the entering tones may have disappeared in some Chinese languages, e.g. Mandarin, where entering tones no longer exist. In addition, this method provides people interested in tones with a rather dependable tool for the study of tones. What I have described here is just the beginning of incorporating some non-linguistic techniques into the field of linguistic study. I hope in the future there will be more use of other computational and statistical techniques in the study of linguistic phenomena.

Notes

1. I am very grateful to Prof. C. C. Cheng, director of the language learning lab, University of Illinois, Urbana-Champaign, who provided me with the computer hardware. I also would like to express my gratitude to Mr. Matthew Ratta, an engineer at Intel Inc., for his help with the K-means algorithm and the computer software programming, to Dorothy Evans and Steve Helmreich for their valuable suggestions and comments, to Rachel Maxwell for her help in the speech recording.

2. Taiwanese is a dialect of Southern Min language. It is spoken by about 10,000,000 people in Taiwan (Zhang 1983).
3. The actual environment for the sandhi tone should be the non-final position in a tonal group. For detailed description, see Chen (1986).
4. The A/D and D/A converter was designed by Data Translation Inc.
5. The reason for using a sampling rate of 10K for male voices is that Rabiner and Schafer (1976: 136) claim that the 10K sampling rate gives a rate sufficient to give adequate time resolution, allowing the period to be determined to within $T = .0001$ seconds. The 20K sampling rate gives a better quality reproduction of the female voice played back through the D/A converter.
6. For these three functions, please refer to the references.
7. Of course, this algorithm only works for monosyllabic words. For poly-syllabic words, the program included a splicing function which could extract a syllable from an input signal and store this syllable speech signal on a disk. Later, this syllable speech signal was loaded and the seven parameters were obtained by the above algorithm.
8. In R. Cheng (1977), there is no falling-rising tone. Our experimental data, however, show that "13" is a falling-rising tone. However, we still use 13 to represent this tone in the computer program and transcription.
9. The normalization equation is $100 \times ((\text{frequency} - \text{lowest frequency}) / (\text{highest frequency} - \text{lowest frequency}))$.
10. The tone identification performance is determined as follows. When we prepared the experiment data, there were 20 monosyllabic words for each tone. Each of the two speakers read 140 words, so there were 40 words for each tone. If each of the seven groups contains only one tone and the number of words with this tone is 40, then the identification performance is equal to the sum of all words in the seven groups divided by the total speech words (280/280). We have a 100% performance, which is perfect. If a group contains words in one tone and the number of words is less than 30, then the performance is not good. This group is dropped. If there are words in different tones included in one group and the number of words in one of the tones is either greater than 10 or less than 30, this group is discarded. Otherwise, the number in each group is added together and the sum is divided by 280. The result is the performance percentage.
11. Chow (1980) claims that a proper-name syllable keeps its citation tone before the diminutive /a/.

References

- Borden, Gloria J. and Katherine S. Harris. 1984. *Speech Science Primer*. Baltimore: Williams & Wilkins.
- Chao, Yuan-ren. 1930. "A System of Tone-Letters". *Le Maître Phonétique*: 24-27.
- Chen, Matthew Y. *The Syntax of Phonology: Xiamen Tone Sandhi*. ms.
- Cheng, Chin-chuan. 1973. *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.
- Cheng, Chin-chuan, and Bruce Sherwood. 1982. "Technical Aspects of Computerassisted Instruction in Chinese". *Studies in Language Learning*, 3.1: 156-170.
- Cheng, Robert L. 1968. "Tone Sandhi in Taiwanese". *Linguistics*, 41: 19-42.
- Cheng, Robert L. 1973. "Some Notes on Taiwanese Tone Sandhi". *Linguistics*, 100: 5-25.

- Cheng, Robert L. and Susie S. Cheng. 1977. *Phonological Structure and Romanization of Taiwanese Hokkien*. Student Co., Ltd. Taipei, Taiwan.
- Chiang, Helen T. 1967. "Amoy-Chinese Tones". *Phonetica*, 17: 100-115.
- Chow, James. 1980. "a Study of a ' ' in Amoy Chinese". ms.
- Ho, Ta-an. 1984. "Two Diachronic Implications of Phonemic Variation, with Discussion on Original Values of Chin-Chiang Tones." *BIHP*, 55.1: 115-132.
- Lieberman, Philip. 1977. *Speech Physiology and Acoustic Phonetics*. Macmillan Publishing Co., Inc.
- Lu, Zhi-ji. 1986. "Tonal Changes: Interplay Between Tone and Tone Sandhi: a Case Study of the Shanghai Dialect". *Studies in the Linguistic Sciences*, 16.1: 97-111.
- Luo, Chang-pei. 1956. *Xiamen Yinyuan Xue*. Science Co., Ltd. Shanghai.
- MacQueen, J. 1965. "Some Methods for Classification and Analysis of Multivariate Observations". *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 281-297.
- Pike, Kenneth. 1948. *Tone Language*. Ann Arbor: the University of Michigan Press.
- Rabiner, L. R. and R. W. Schafer. 1978. *Digital Processing of Speech Signals*. New Jersey: Prentice-Hall, Inc.
- Sondhi, M. M. 1968. "New Methods of Pitch Extraction". *IEEE Trans. Audio and Electroacoustics*, AU-16.2: 262-266.
- Tseng, Chiu-yu. 1981. *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Ph.D dissertation, University Microfilms International, Ann Arbor, Michigan.
- Tung, Chao-hui. 1964. "Phonology of Taiwanese as Spoken in the Kaohsiung Area". *Bulletin of Taiwan Normal University*, 9: 195-204.
- Tung, T'ung-ho. 1957. "Phonology of the Amoy Dialect". *BIHP*, 29: 231-253.
- Tung, T'ung-ho, Yung-lang Chao and Ya-shiu Lan. 1967. *A South Min Dialect of Taiwan*. BIHP, Monographs Series A. No. 24. Taipei, Taiwan.
- Wang, William S-Y. 1967. "Phonological Features of Tone". *International Journal of American Linguistics*, 33: 93-105.
- Weingartner, Fredric F. 1970. *Tones in Taiwanese: an Instrumental Investigation*. National Taiwan University, College of Arts. Monograph No. 2.
- Yip, Moira Jean Winsland. 1980. *The Tonal Phonology of Chinese*. Ph.D dissertation, MIT. Reproduced by the Indiana University Linguistics Club.
- Zhang, Zhen-xing. 1983. *Taiwan Minnan Fanyian Chi Lueh*. Foochow.