A Study of Speaker Identification : an analysis and perception of sisters′ speech

Kim Kyung Hwa
(The Supreme Public Prosecutor's Office, Korea)

1. INTRODUCTION

Human voice, like your fingerprints, is one of distinctive features that represents a speaker's identity. However, there is great similarity of voices between family members, such as sisters, brothers, and twins as well, and it is often very confusing to distinguish them, especially when they are heard through a telephone.

There have been many studies dealing with voice identification, but a few of them were performed on family members who are thought to have similar voices. In this study, therefore, the prime target will be put on the analysis of sisters′ speech, as a pilot study.

Two kinds of experimentation have been conducted for this study: a visual examination of spectrograms and an aural discrimination test. In the first one, mean frequencies of formants F3, F4 of selected vowels and F0 were measured, then compared with each other, by pairs. In the second, by the aural test, it was observed whether the listeners distinguished sisters′ voices, one from the other, and what the clues are for recognition of voice differences.

2. Previous Studies

Concerning with this study, there are several studies on twins′ speech.

Lundstrom(1948) indicated that the factors governing the variation in identical pairs of twins are of similar kind as those causing differences between the halves of the body. [1]

Nolan & Oh(1996) examined differences in articulation of the phonemes /r/ & /l/ by identical twins. Keith Johnson & Misty Azara(1997) found out that identical twins do not talk identically so listeners were pretty good at telling the twins apart from each other. [2] [3]

3. Experiment 1

3.1. Subjects

The subjects were 10 females consisting of 4 pairs of sisters (2×2pairs, 3×2pairs). All of them are from Seoul and aged from mid 20's to late 30's. The speakers' initials are KKH, KDY / YSJ, YSY / SJK, SMK, SYK / KOJ, KSJ, KMJ

Materials

The materials consist of various types of speech such as isolated words with carrier sentences, casual speech, and reading sentences, to extract more speaker' specific features. They are 16 isolated words, 17 casual speech, 16 reading sentences. Each speaker read them 3 times.

For the method of visual examination of spectrograms, Tosi et al., (1972) considered mean frequencies and bandwidth of vowel formants, gaps and types of vertical striations, slopes and transients of formants, and energy distribution of fricatives and plosives etc. Atal(1972) selected pitch contours as optimal parameters For a particular method of automatic talker recognition. Wolf(1972) selected fundamental frequency at given locations of the sample sentences, amplitude of filtered vowels, mean frequencies of formants F1 and F2 in given locations of the sample sentences etc. [4] [5] [6]

We measured mean frequencies of formants F3 and F4 of 'a, e, i, o, u' in sample words and F0 of the first syllable of reading sentences.

3.2. Procedure

All recordings were made by DAT TCD-D7 and Computerized Speech Lab. For analysis, using the method of FFT in addition to LPC in PitchWorks, frequencies of formants of vowels in their stable areas were measured. And then their value were evaluated by the naked eye.

Among the total 1,470 sentence readings (49 samples x 3 repetitions x 10 speakers), 150 (5 samples x 3 repetitions x 10 speakers) were analyzed for F3 and F4, and 39 (13 samples x 3 repetitions x 10 speakers) for F0, excluding syllables initiating with ㅋ, ㅎ, ㅆ that have rather high pitch.

3.3. Results

3.3.1. Mean frequencies of formants F3, F4 of vowels

The mean frequency of formants F3 and F4 of vowels 'a, e, i, u' showed similarity in some sisters, but not in all, and the degree of their similarity differed by sisters and vowels. The table 1 and 2 show mean frequencies of F3 and F4 (the empty cell on the table means that formants showed no value on the spectrogram, and vowel 'o' was excluded in this analysis because F3 and F4 of it could not be observed in more than half of speakers' speeches).

<Table 1> Mean frequencies of F3, F4 (Hz)
(KDY, KKH / SJK, SMK, SYK)

	KDY	KKH	SJK	SMK	SYK
'a'F_3	1834	2091	2195	2143	2391
SD	99.43	56.01	9.50	127.87	107.85
'a'F_4	3065	3074	3285	3249	3421
SD	35.47	78.01	76.21	48.05	67.30
'e'F_3	2481	2872	3172	3209	3135
SD	146.77	254.13	16.44	17.16	84.55
'e'F_4	4147	4359	5044	4331	4357
SD	22.05	6.00	141.78	48.51	185.27
'i'F_3	2947	3266	3369	2987	3473
SD	57.12	134.51	87.83	82.56	108.26
'i'F_4	4187	4229	4342	4271	4509
SD	57.62	47.51	103.16	124.04	80.53
'u'F_3	1727	1964	1706	1762	1665
SD	63.96	276.10	76.51	181.51	191.23
'u'F_4	2553	2703	3023		2829
SD	17.04	50.93	4.73		50.92

<Table 2> Mean frequencies of F3, F4 (Hz)
(YSJ, YSY / KOJ, KSJ, KMJ)

	YSJ	YSY	KOJ	KSJ	KMJ
'a'F_3	2200	2125	2095	2449	2062
SD	140.88	59.80	133.98	96.02	225.62
'a'F_4	3331	3341	3266	3424	3098
SD	101.3524	131.07	83.23	71.58	55.19
'e'F_3	3273	3206	3036	3210	2921
SD	65.3835	30.62	106.52	27.43	24.21
'e'F_4	4610	4621	4343	4435	4135
SD	96.2497	42.46	84.72	45.76	131.04
'i'F_3	3266		3172	3523	2870
SD	69.3974		73.57	84.24	29.02
'i'F_4	4584	4632	4137	4385	4095
SD	56.3116	89.96	195.48	123.88	52.62
'u'F_3	1881	1833	1715	1630	1539
SD	206.3234	70.79	153.64	130.33	158.21
'u'F_4	2818	2788	2723	2991	2671
SD	57.0731	26.00	102.59	94.69	69.66

Sisters who showed the most similar frequency of formants were the case of YSJ-YSY. They were almost identical in the value of vowels 'a, e, u'. And some sisters who showed a difference in absolute figures of mean frequencies of formants showed similar gaps between F3 and F4 in their speeches.

Among the words analysed, one sample word was used to observe the change of the value of the formants of the same speaker, and, in order to check the recognition rate when they were heard, each speaker was asked to say twice in normal speech, and in disguised way for the third. The speakers intentionally changed their voices, in low or high pitch, but it didn't affect much the mean frequency of formants.

Some of previous studies conducted research into the variation of formants based on the change of vocal pitch, and decided the intensity of high formants was increased in high pitch. [7]

Table 3 shows each speaker's average F0. (KOJ was excluded because she didn't give pauses between sentence numbers and the first syllables of sentences)

<Table 3> average F0 of each speaker

	average F0 (Hz)	SD
KDY	212.8	19.0
KKH	214.5	13.2
YSJ	215.3	12.4
YSY	202.2	10.9
SJK	213.1	16.1
SMK	205.6	12.4
SYK	232.5	18.9
KSJ	224.0	7.3
KMJ	224.2	13.5

Sisters KDY_KKH and KMJ_KSJ have almost same F0, but the other 2 sisters showed a little different F0.
For the present, there are no decisive criteria to decide how much difference of the F0 could be said similar, or different.

4. Experiment 2 - Aural test

For the second experiment, aural tests were performed to examine whether or not the listeners could tell the difference between the two or among the three sisters, and to find out what made them able to distinguish unknown two voices.

4.1. Materials

In experiment 2, 40 paired samples from among those used in experiment 1 were selected which consist of 15 isolated words, 14 casual speech, and 11 reading sentences. And mixed pairs spoken by the same talker, sisters, and unrelated talkers were given to the listeners, and let them tell whether the voices were from the same speaker or not. Listeners were not told that talkers consisted of 4 different groups of sisters, and that their changed voices were included in the given materials. Listeners participated in the experiment were 16 from the Supreme Public Prosecutor's Office: 13 males and 3 females.

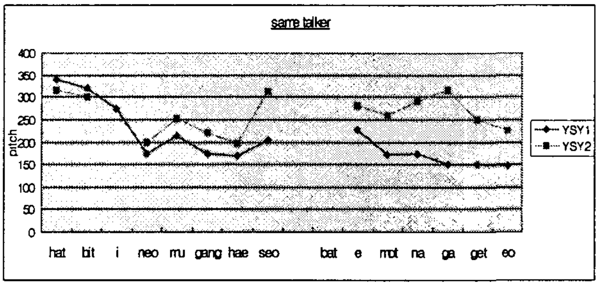
4.2. Results

The result shows that average 70.2 % of the listeners gave correct answers. The most undistinguished word was "metuki" spoken by the sisters YSJ-YSY, with only 12.5 % correct answers. When the value of formants of vowels from "metuki" told by them, it was found out that the locations of the formants were similar.

Recognition rate for each speaker ranged between 56.2% and 100%, and listeners showed a strong tendency that they took the same speaker for different speakers, expecially when short utterances or utterances spoken in disguised voices were given to them. Table 4 shows the recognition rate of each talker and Figure 1 is an example of confusing pairs.

<Table 4> The recognition rate

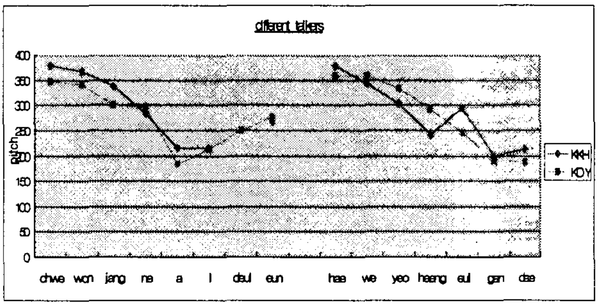
Talker	recognition rate(%)
KDY	100.0
KKH	96.9
SJK	56.2
SMK	64.6
SYK	81.2
YSJ	77.1
YSJ	75.0
KOJ	93.7
KSJ	75.0
KMJ	90.6



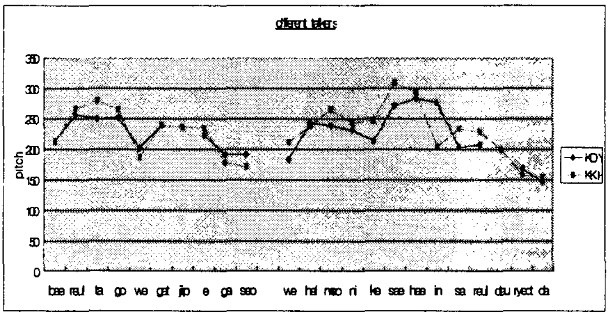
<Fig. 1> the case which many listeners regard same talker as different talkers. (F0 : YSY1 - 207, YSY2 - 263)

When the groups of sisters are compared, the pair of KSJ-KMJ was marked the highest rate of correct answers(90.6%), and the other three groups' were below 50 percent: 41.6% for YSJ-YSY, 42.5 % for SJK-SMK-SYK, and 45.8% for KKH-KDY. In comparison among unrelated pairs of speakers, the rate of recognition was rather low, 50% for SJK-KSJ and 56.2% for YSJ-KOJ, and both made the listeners confusing. The former pair talked in low pitch as disguised voices, and the latter talked in a similar speech style of sweet and soft.

One interesting fact was that, in case of the pair of sisters KKH_KDY, the recognition rate for each speaker reached almost 100%, but when mixed pairs of their talks were given, the rate dropped to below 50%. This phenomenon appeared in both long and short speeches. When their pitch contours of each syllable in the given sentences were measured and compared each other, it was found that the pitch contours for two speakers were very similar. Figure 2 and 3 show it.



<Fig. 2> the case which many listeners regard different talkers as same talker. (F0 : KDY - 288, KKH - 278)



<Fig. 3> the case which most listeners regard them as same talker.
(F0 : KDY - 223, KKH - 229)

According to this experiment, key factors that make listeners take different speakers for the same one, or the same for different ones probably are the pitch and speech style, not the difference of pronunciation or speech rate.

5. CONCLUSION

In this study, speeches of four groups of sisters were analyzed, individually and by group, and then compared with each other, to observe their characteristic and similarity. At first, measurement was made on mean frequencies of formants F3 and F4 of vowels 'a, e, i, u' and F0 of their speeches. Then, through aural tests, recognition rate for the target sisters' speeches was observed, and the need of criteria for voice identification was discussed.

As a result, the formants F3 and F4 of sisters', in some cases, show similarity in location and in gap between F3 and F4, but, in general, this similarity was not enough to be defined as a characteristic of sisters.

The question about how to interpret the difference of formants has close relationship with the question about how much difference of formants can be taken for the same speaker's, and those questions would be answered after serious discussions and a lot more experiments are performed.

Two of the four groups of sisters were observed to have similar value of F0. But the question is the range of difference of F0 that will decide speeches as one or more speakers'. More intensive and continuous studies should be conducted to provide its criteria.

Though not involved in this study, for speaker recognition, various factors mentioned above should be put into consideration: such as slopes and transients of formants of F1 and F2, and the gap between them, and energy distribution of fricatives and plosives, etc. Through these works, general characteristics of sisters' voices could be found. At the same time, long-term research and study should be peformed on intraspeaker variability.

REFERENCES

[1] Lundstrom, Anders, Tooth Size and Occlusion in Twins. S. Karger, Basle, 1948.

[2] Nolan & Oh, "Identical Twins, Different Voices". Forensic Linguistics 3:39-49, 1996.

[3] Johnson, Keith. & Azara, "The perception of personal identity in speech: Evidence from the perception of twins speech", Acoustic Society of America, 1997.

[4] Tosi, Oscar. et al., "Experimental on voice identification", Journal of the Acoustic Society of America, 51:2030-2043, 1972.

[5] Atal, B. "Automatic speaker recognition based on pitch contour", Journal of the Acoustic Society of America 52:1687-1697, 1972.

[6] Wolf, J., "Efficient acoustic parameters For speaker recognition", Journal of the Acoustic Society of America 51:2044-2056, 1972.

[7] 유영화 외 3인, “Pitch의 변화가 Formant에 미치는 영향에 관한 연구”, 국립과학수사연구소연보 제 18권, Vol 18, 1986.