Shape Display: Task Design and Corpus Collection

Janice Fon

Graduate Institute of Linguistics
National Taiwan University
jfon@ntu.edu.tw

Abstract

This study introduces a new paradigm for spontaneous dialog elicitation and a small multilingual corpus collected using this paradigm. Pairs of subjects were seated in separate booths and were each given a felt-covered board and a bag of assorted felt pieces of various shapes and colors. The goal was to make the layout of the felt pieces the same on the two boards with the least moves. In order to test how accommodating the paradigm is to cross-linguistic/cross-cultural experimental designs, 32 subjects of three different languages, English, Mandarin (Guoyu and Putonghua), and Japanese participated in the study. Subjects found the paradigm entertaining and engaged themselves in the game without paying much conscious attention to their linguistic performances. The elicited dialogs were spontaneous enough to allow further phonetic and discourse research.

1. Introduction

Although traditionally, phoneticians feel more comfortable working with controlled rather than spontaneous data, more and more studies have started to investigate more natural renditions of speech. This trend is based on the now common belief that scripted and unscripted speech types are two distinct genres and what applies to the former does not necessarily apply to the latter. Read speech, although more controlled and easier to handle, can never parallel the richness and the dynamics of spontaneous speech, a genre we encounter almost everyday in our daily lives.

However, unlike in other subfields of linguistics, where a certain level of acceptable spontaneity can be easily obtained from subjects by simply hiding a walkman in an inconspicuous corner in places where the subjects are comfortable with, phonetic studies have their natural constraints. In order to facilitate further phonetic/acoustic analyses, speech scientists demand high quality of recording. As a consequence, the recording setting is usually in a sound-treated room that is unfamiliar and oftentimes intimidating to the speakers and the high-fidelity recording equipment hooked onto the speakers could only serve as a flamboyant (and sometimes even uncomfortable) reminder of the unnaturalness of the setting. In other words, the requirement for a clean recording often clashes with that for a natural setting to elicit spontaneous speech.

In order to solve this “inborn” dilemma, various paradigms for eliciting dialogs have been devised, all of which were intended to make the unnatural setting somewhat more natural. The most famous of all is the Map Task developed by the Human Communication Research Centre in Britain [1].

In the Map Task, there are two speakers, one is designated as the instruction giver and the other the instruction follower. Both are given maps but the two maps are not exactly identical. The map for the instruction giver has a route marked while that for the instruction follower does not. The goal is for the pair to cooperate so that the instruction follower reproduces the route. Corpora using the Map Task have been built in many languages including English (e.g., [2]), German (e.g., [3]), and Japanese (e.g., [4]).

However, there are some constraints regarding the kind of dialogues elicited using the Map Task. Because of the task design, the roles assigned to the two participants of a pair are not equal. The instruction giver is the major contributor of the dialogue while the instruction follower plays a more minor role. Though this type of unequal dialogues is important and practical on many occasions (e.g., at an information desk), it is also interesting and vital, both theoretically and applicationwise, to study dialogues of which participants are more equal in contribution, which is also common in our everyday lives (e.g., at a negotiation table).

Another constraint of the Map Task lies in its cultural dependency. Many of the landmark icons on the maps rely heavily on cultural contexts, and it is sometimes difficult to name them correctly without referring to the typed caption below. This would become a more serious problem when designing comparable corpora for a variety of languages, as it is then essential to have exactly the same experiment setup and elicitation paradigm.

This study reports a new paradigm, the Shape Display, for spontaneous dialogue elicitation. The Shape Display is designed with three specific goals in mind to accommodate the needs for employing phonetic analyses on natural dialogues. The first goal is spontaneity. As mentioned before, due to the constraints of phonetic studies, it is essential to create an experimental setting that is as natural and as engaging as possible so as to divert participants from paying conscious attention to their linguistic performances. The second goal is equal dialogue status. This is one of the main purposes in creating a new dialogue elicitation paradigm in addition to the already existing Map Task. The third is to devise a paradigm that is more culturally independent so as to facilitate the construction of multilingual corpora.

2. Task design

This section introduces the detailed layout of the paradigm, including the number of participants, the equipment needed, the experimental setting, and the procedure.

2.1. Participants

Two participants are required to complete this task, although theoretically speaking, this task can accommodate more than two if situation permits.

2.2. Equipment

Three pieces of equipment are needed in a Shape Display game, including the display board, the shape pieces, and the

shape pocket. To minimize shuffling noise, all except for the board is made of felt. The board is also covered with felt.

Two display boards of some maneuverable size (approximately 42 cm × 27 cm) are needed. The surface is divided into six equal sections by pinning threads of yarn on the board, as shown in Figure 1.

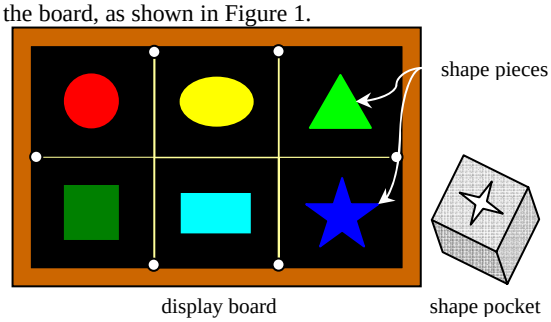

Figure 1: *The equipment of the Shape Display game, a display board, 36 shape pieces, and a shape pocket.*

There are in total six different shapes of shape pieces, including circle, oval, triangle, square, rectangle, and star. The inventory does not have to be limited to these six, but can be modified accordingly to suit individual needs. The shapes here are chosen because they have a relatively high degree of cultural independence and are suitable for establishing a multilingual corpus. Shape pieces also vary in color. Currently, six colors are included. They are red, yellow, light green, dark green, light blue, and dark blue. The colors are chosen because they are relatively common color terms in languages. Two shades of green and blue are included so that possible focus contrasts can be elicited. In total, there are 6 (shapes) × 6 (colors) = 36 shape pieces for each person (Figure 1). Of course, the number of shape pieces does not have to be limited to 36, but can be modified according to individual experimental needs. However, more pieces usually entail longer game time and sometimes participants might thus be discouraged due to consecutive unsuccessful trials. The shape pieces are mixed well and put into a sewn felt pocket. The pocket is shaped as a cube and has an opening slit on the top. The slit is small enough so that one cannot see through and tell what is inside. Each participant has one felt pocket (Figure 1).

2.3. Experimental setup

The recording should be done in a quiet room with two separate sound booths. There should be at least one chair and one desk in each booth. The subjects should be seated in different booths so that they could not see each other. To avoid uneasiness with the recording equipment and obtain clean recordings, head-mounted microphones should be used. Recordings are made so that one subject's voice goes into the right channel of the recorder and the other's goes into the left channel.

2.4. Procedure

Before the recording, the display boards and the shape pockets are placed inside the sound booths and the experimenter randomly picks six shape pieces out of each shape pocket and places them on each board in random order. The subjects are explained the rules of the Shape Display game before they enter the booths. They are told that there are six shape pieces on each of the display board, and the goal of the game is to make the display on the two boards look exactly the same. They can do so by switching positions of existing shape pieces on the boards or drawing new pieces out of the shape pocket. The unwanted pieces should not be put back into the pocket but should be placed on the side. They could achieve the goal by using whatever strategy they prefer but they should try to minimize the number of pieces they use throughout the game. The one that uses the least pieces "wins" the game.

3. Corpus collection

In order to test whether the Shape Display paradigm is useful in eliciting equal spontaneous dialogs from various cultures and languages, a small multilingual corpus was collected. The collected languages include English, Mandarin, and Japanese. Two dialects of Mandarin, Guoyu and Putonghua, were included in the corpus. The former is the national language of Taiwan and the latter of Mainland China.

3.1. Subjects

Subjects from three languages, English, Japanese, and Mandarin, and two dialects of Mandarin, Guoyu and Putonghua, were recruited. All speakers were recruited from Columbus, Ohio. Japanese and Mandarin native speakers were recruited through student associations and personal connections. In order to attain homogeneity, only native speakers of Central Ohio English from Columbus and neighboring counties [5], Tokyo Japanese from Tokyo and three neighboring prefectures (i.e., Chiba, Saitama, and Kanagawa), Taipei Guoyu from Taipei City and Taipei County, and Beijing Putonghua from Beijing area were included. Subjects were either born and raised in the language area or moved there before age three, and had no exposure to other languages before that. It was very difficult to find pure monolingual speakers for all three languages. English speakers from the local community often took classes in foreign languages during their high school and college years to fulfill the second language requirement. Of the eight English subjects recruited, only two claimed that they do not know any language other than English. For Mandarin and Japanese speakers, it was even more unlikely to record monolingual speakers, not only because they were recruited from the States, but all speakers in the two language groups were required to take English as their second language in elementary and/or high school. However, Japanese and Mandarin speakers that lived in the States for more than three years, and had not used their native languages as the dominant everyday language since moving to the U.S. were excluded. Four female and four male subjects were recruited for each language/dialect group. Thus, there were 4 (subjects) × 2 (genders) × 4 (languages/dialects) = 32 subjects in total. Subjects were paid with monetary rewards. Table 1 shows the demographics of the subjects recruited.

Table 1: *Demographics of recruited subjects.*

	English	Guoyu	Putonghua	Japanese
X	25.25	28.25	29.125	26.5
SD	6.36	1.48	3.72	5.52

3.2. Equipment

Due to the limited number of head-mounted microphones with headphones of the same type, the two subjects of a pair used different types of microphones. One used a SHURE SM10A head-mounted microphone coupled with SONY MDR-7502 headphones. The other used a Sennheiser HMD25-1 head-mounted microphone with headphones. The subject that had a bigger head of the two was designated to use the former. Both microphones and headphones were connected to a SONY DAT DTC-790 recorder through a Symetrix SX202 Dual Mic Preamp preamplifier. Maxell R-64DA 60 min DAT tapes were used for the recording.

3.3. Procedure

Recordings were made in the Phonetics Laboratory in the Department of Linguistics, The Ohio State University. The procedure generally followed the Shape Display paradigm described above. To avoid any accommodation effect that subjects might have, the recording for each group was done by an experimenter that was a native or a near-native speaker of the dialect. As a native speaker of Guoyu, I did the recording for the group and asked four other experimenters to help with the recordings for the other language groups.[1] Each of the subjects did two recordings, one with a partner of the same sex, and the other with that of different sex. In total, there were 4 (pairs) × 2 (trials) × 4 (languages/dialects) = 32 dialogues. The recording duration of each subject pair is shown in Table 2. In total, the corpus contained over 190 min of recording, of which approximately 38 min was English, 44 min was Guoyu, 53 min was Putonghua, and 56 min was Japanese. A D-to-D transfer was then done at a sampling rate of 44100 Hz and the files were later downsampled to 22050 Hz for further analyses.

Table 2: *Recording duration of subject pairs.*

English		Guoyu	
ID	Duration	ID	Duration
AL+JH	3:10.295	CHL+HPH	4:12.133
AL+MD	6:00.019	CHL+SMI	5:18.890
BH+DF	5:14.548	CYW+HPH	6:47.261
BH+PA	4:08.018	CYW+SMI	8:02.959
MK+JH	4:23.813	HHY+SCY	5:35.791
MK+MD	5:16.345	HHY+WCF	4:15.211
SU+PA	5:13.916	SCH+WCF	5:46.506
SU+DF	4:26.270	SCH+SCY	3:45.505
Total	37:53.224	Total	43:44.256
Putonghua		Japanese	
ID	Duration	ID	Duration
CY+FBS	5:11.618	FY+SA	7:10.410
CY+LX	8:07.230	FY+MS	7:10.349
LJG+FBS	4:32.826	NY+IA	7:06.888
LJG+LX	5:53.098	NY+UK	6:59.867
YYL+XD	6:34.789	NT+IA	4:53.269
YYL+ZLE	12:15.563	NT+UK	6:34.653
ZLI+XD	4:58.746	YE+MS	7:29.949
ZLI+ZLE	5:35.841	YE+SA	8:54.597
Total	53:09.711	Total	56:19.989

[1] Two experimenters recorded for the English group because the first experimenter was not available for some of the later recordings.

3.4. Transcription and discourse labeling

Recordings were orthographically-transcribed by (near-) native speakers. In total, approximately 5500 words were collected for English, 6300 words for Guoyu, 8000 words for Putonghua, and 8300 words for Japanese.

Besides basic transcription, a prosodic transcription based on the Tone and Break Indices (ToBI) systems of respective languages using Praat is also planned (English: [6]; Mandarin: [7] & [8]; Japanese: [9]). In addition, a discourse labeling tier based on Grosz & Sidner's model [10] following the labeling principles in my dissertation [11] will be created. Also, since the current corpus is of dialogues, a separate tier coding the turn-taking situation between the two speakers will be labeled. An example of the labeling scheme is shown in Figure 2. The two waveforms represent two speakers in a dialogue. Currently, there are three ToBI tiers for each speaker ([7] & [11]). The top is the word tier which codes word labels, the middle is the syllable tier, and the bottom is the miscellaneous tier.

3.5. Discussion

Looking at the corpus collected, one finds that the Shape Display paradigm is successful in eliciting spontaneous dialogues. The three preset goals were reached to a satisfactory degree. In terms of spontaneity, the use of language was fairly colloquial, and the style was casual. Subjects were much engaged in the game. A brief informal interview with them after the recording showed that in general, they found the game entertaining and had fun, which is one of the essential elements for obtaining natural speech. In terms of equal dialogue elicitation, one also finds that the corpus contains frequent exchanges and in most cases, participants take turns in leading the dialogue. Finally, the paradigm also seems to work well with speakers from different languages and cultures. Except for the medium of instruction, no other modification is needed across language groups.

4. Conclusions

This study introduces a new paradigm, the Shape Display, for spontaneous dialogue elicitation. The paradigm requires a pair of subjects to play a simple board game. Competition is involved to motivate subjects and divert them from paying conscious attention to their linguistic performances. A small multilingual corpus was collected using the paradigm to test whether it is reliable for eliciting natural dialogues. Results showed that the paradigm could successfully elicit spontaneous dialogues in which both participants contribute more or less equally. More importantly, the multilingual corpus, although small, also showed that such a paradigm is useful in eliciting corpora cross-linguistically. Except for the medium of instruction, which should be in respective languages, no other modification is needed.

5. Acknowledgements

The design of the corpus was supported by a subproject under the 1998-1999 interdisciplinary seed grant on Spoken Language Understanding and Generation, funded by The Ohio State University Office of Research (PIs: Mary Beckman, Marjorie Chan, Robert Kasper, Terrell Morgan, Craige Roberts, and Don Winford). The collection of the corpus was supported by a research grant funded by the

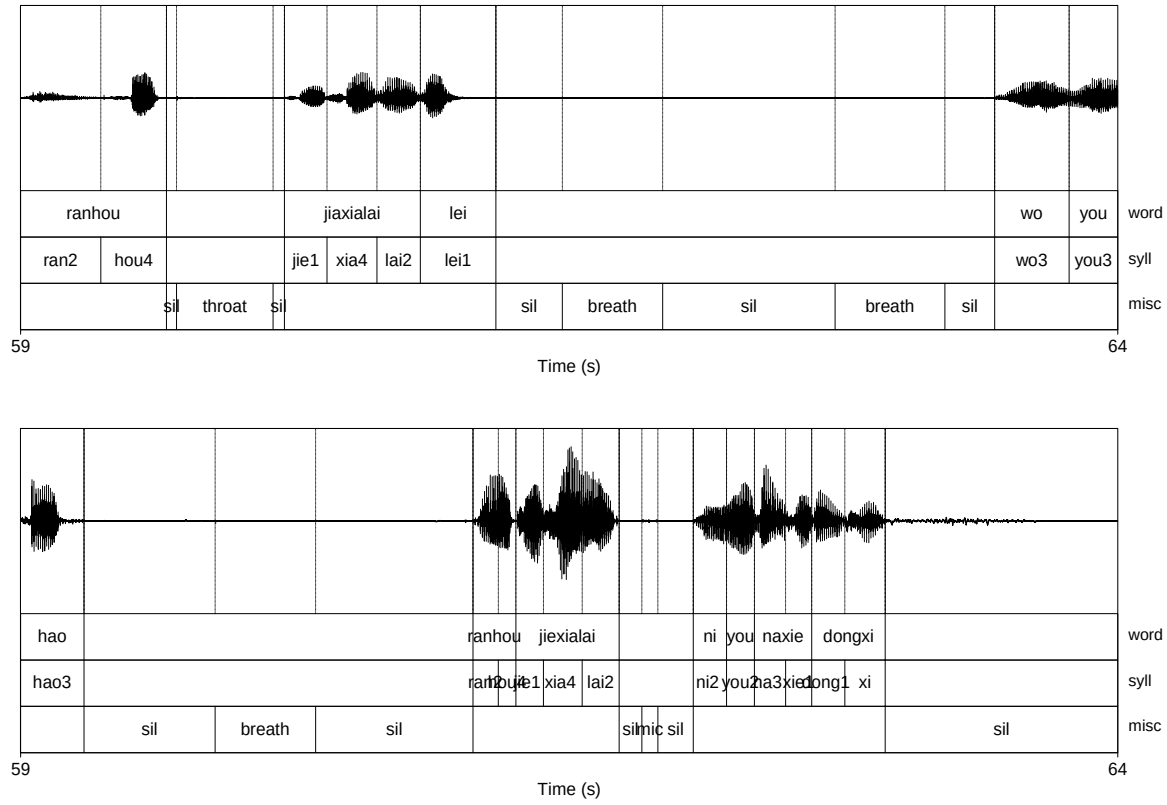

Figure 2: *A Guoyu example of the current labeling tiers.*

National Institute of Development and Communication Disorders (grant no. R01DC04421) awarded to my mentor and advisor, Keith Johnson, to whom I am deeply thankful. I would like to express thanks to my experimenters Keith Johnson (English), Terah Schamberg (English), Tsan Huang (Putonghua), and Kiyoko Yoneyama (Japanese). Many thanks also to my transcribers Terah Schamberg (English), Tsan Huang (Putonghua), and Aya Shirai (Japanese). Without them, the experiments could not have been run so smoothly. Naturally, all the faults are mine.

6. References

[1] Anderson, A.; Bader, M.; Bard, E.; Boyle, E.; Doherty, G. M.; Garrod, S. et al., 1991. The HCRC Map Task Corpus, *Language and Speech*, 34, 351-366.

[2] Lickley, R.; Bard, E., 1998. When can listeners detect disfluency in spontaneous speech?, *Language and Speech*, 41(2), 203-227.

[3] Mixdorf, H., 2004. Qualitative analysis of prosody in task-oriented dialogs, *Proceedings of the 2nd International Conference on Speech Prosody*, 283-286.

[4] Koiso, H.; Horiuchi, Y.; Tutiya, S.; Ichikawa, A.; Den, Y., 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs, *Language and Speech*, 41(3/4), 295-321.

[5] Flanigan, B. O; Norris, F. P., 2000. Cross-dialectal comprehension as evidence for boundary mapping: Perceptions of the speech of southeastern Ohio, *Language Variation and Change*, 12, 175-201.

[6] Beckman, M. E.; Ayers Elam, G., 1997. *Guidelines for ToBI labelling.* Unpublished manuscript, The Ohio State University.

[7] Peng, S.-H.; Chan, M. K. M.; Tseng, C.-Y.; Huang, T.; Lee, O. J.; Beckman, M., 2000. *A pan-Mandarin ToBI.* Unpublished manuscript, The Ohio State University.

[8] Tseng, C.-Y.; Chou, F.-C., 1999. Machine readable phonetic transcription system for Chinese dialects spoken in Taiwan. *The Journal of the Acoustical Society of Japan (E)*, 20(3), 215-223.

[9] Venditti, J. J., 1997. Japanese ToBI labelling guidelines. In *The Ohio State University Working Papers in Linguistics*, K. Ainsworth-Darnell & M. D'Imperio (eds.), 50, 127-162.

[10] Grosz, B.; Sidner, C. L., 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175-204.

[11] Fon, Y.-J. J., 2002. A Cross-linguistic Study on Syntactic and Discourse Boundary Cues in Spontaneous Speech. Dissertation. The Ohio State University.