論文 / 著書情報
Article / Book Information

Title	A Robot ' s Slip of the Tongue: Effect of Speech Error on the Familiarity of a Humanoid Robot
Author	Takayuki Gompei, Hiroyuki Umemuro
Journal/Book name	Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication, , , p. 331-336
Issue date	2015, 8
DOI	http://dx.doi.org/10.1109/ROMAN.2015.7333630
URL	http://www.ieee.org/index.html
Copyright	(c)2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.
Note	このファイルは著者（最終）版です。 This file is author (final) version.

A Robot's Slip of the Tongue: Effect of Speech Error on the Familiarity of a Humanoid Robot

Takayuki Gompei[1] and Hiroyuki Umemuro[1]

Abstract— **This study investigated the effect of speech errors made by a robot on people's perception of the familiarity of the robot. Four types of speech errors were implemented for a communicating robot. Subjects' perception of the familiarity of the robot, along with other subjective evaluations of impressions of the robot, were measured in an experiment and compared between conditions with and without speech error. Results showed that speech error improved the familiarity score when the robot did not make any speech errors in the first contact with subjects and then made speech errors in the second contact on a separate day. On the other hand, speech errors lowered people's perception of the sincerity of the robot. These results imply that the robot should not make speech errors in the early stage of engagement with human users, while some speech errors after the users become accustomed with the robot might be effective in improving users' perception of the familiarity of the robot. Results also showed that speech error improved familiarity for people having low-scoring attitudes about the diversity and operation of robots.**

I. INTRODUCTION

A sense of familiarity is essential to people's acceptance of and engagement with robots. People who feel familiar with robots tend to wish to communicate with them[1], and consequently, they might feel that the robots are more useful. It is thus essential that robots engage with people and thus provide people with a greater feeling of familiarity.

The humanlike quality of robots is essential in inducing a feeling of familiarity. Okada [2] noted that for users to feel familiar with robots, the robots should not induce fear in people, and the robots should thus have human-likeness.

Among possible approaches for developing the human-likeness of robots, one approach that has attracted attention is the control of non-verbal information[3]. Non-verbal information can refer to gestures, prosody, and the speech act. The speech act, or performative utterance[4], refers to certain actions that are performed while a person is saying something. The speech act may include persuading someone, enlightening someone, or getting someone to do or realize something, whether intended or not.

One kind of speech act that is common among adults is speech error. Speech error is a non-intentional slip relative to the utterance intended by the speaker, and includes addition, drop, substitution, swap, contamination, and transfer errors[5]. For human beings, various effects of speech error have been studied, and speech error is known as a factor contributing to humanness[6].

[1]Takayuki Gompei and Hiroyuki Umemuro are with Graduate School of Decision Science and Technology, Tokyo Institute of Technology, 152-8552 Tokyo, JAPAN {gompei.t.aa, umemuro.h.aa}@m.titech.ac.jp

Sakamoto and colleagues[3] implemented falter in speech for a communication robot, and found that people wished to continue interacting with the robot more when the robot's speech faltered. This result implies that implementing human-like speech acts may contribute to people's perception of familiarity with robots. However, the effect of speech error, which is a different kind of speech act from faltering, has not yet been investigated.

The present study investigates the effect of speech errors of a robot on people's perception of familiarity with the robot while they are communicating with the robot. To this end, four types of speech errors were implemented for a humanoid robot. Subjects' perception of familiarity, along with other subjective evaluations of impressions, were measured and compared between the cases of the robot communicating with speech errors and the robot communicating without speech errors.

Additionally, people's attitudes towards robots may affect their perceptions of robots[7]. This implies that the subjective evaluation of the impression of robots may be affected by people's attitudes towards robots. Thus, the present study also assessed people's attitudes towards robots and compared the effects of speech errors on impressions between groups of subjects having positive and negative attitudes in several attitude dimensions.

II. METHOD

A. Experiment Design

A two (speech error: with and without) by two (time of speech error: on the first day versus second day) by two (order of conversation contents: content 1 versus content 2 on the first day) by two (sex) factorial design was employed in this study. Speech error was a within-subject variable, while the other factors were between-subject variables.

B. Subjects

The subjects were 44 undergraduate and graduate students, aged between 18 and 25 years (M = 21.5 years, SD = 1.4 years), attending a science and engineering university in Japan. Of the subjects, 28 were male and 16 were female. Seven males and four female subjects were assigned to each of four condition groups. All subjects were Japanese and spoke Japanese as their native language.

C. Apparatus

NAO[8] (NAOqi OS 2.1) was employed as the platform for the implementation of the conversation with subjects, and conversations were implemented using Choregraphe development environment [9].

Four different contents of conversations were prepared. Two were for training sessions held on the first and second days, and consisted of a few greeting interactions. These sessions each took less than a minute. The remaining two conversations were used in the main measurement sessions of the experiment. Each consisted of a short conversation with several interactions, lasting 2 to 3 minutes. So that the two contents were as similar to each other as possible, both related to trivia on food relevant to the season: a hot pot (content 1) and stew (content 2). As all subjects spoke Japanese as their native language, all conversations were prepared in the Japanese language.

Among the six types of speech errors[5] discussed in the previous section, addition, drop, substitution, and swap errors were chosen based on their frequency of occurrence in human conversations as reported in the literature[5] and their naturalness in the conversations prepared for this experiment. All four types of errors were implemented in each of the two main conversations under the with-speech-error condition. Each type of speech error occurred once in a main conversation session (content 1 or 2), and thus, each conversation in with-speech-error condition had four speech errors. Speech errors were implemented by programming Choregraphe[9] so that the pronunciations of the corresponding parts of the conversations could be heard as natural speech errors.

Table I is a sample of conversations used in the experiment.

D. Procedure

Subjects participated in the experiment one at a time. Each subject participated in two sessions held on two separate days. The session for each day took about 15 minutes per subject.

On the first day of the experiment, the outline of the experiment was first explained to the subjects and the subjects were then asked to complete a consent form if they agreed to participate.

Subjects were then asked to complete the first questionnaire. The first questionnaire asked for demographic information and assessed the subjects' attitudes towards robots.

After completing the first questionnaire, subjects undertook a training conversion session so as to become accustomed with conversing with the robot. The subjects then participated in one main conversation with the robot that was assigned to the subject.

To assess and eliminate potential effects of the order and contents of conversion, four combinations of factors were prepared.

- Day 1: content 1, with errors; Day 2: content 2, without errors.
- Day 1: content 1, without errors; Day 2: content 2, with errors.
- Day 1: content 2, with errors; Day 2: content 1, without errors.
- Day 1: content 2, without errors; Day 2: content 1, with errors.

Each subject was assigned to one of the four combinations of conditions, and participated in the conversion of the assigned conditions.

Finally, subjects were asked to complete the second questionnaire. The second questionnaire comprised two measurements; one assessed the subjects' impressions of the robot while the other assessed the subjects' trust in the robot.

The session on the second day was the same as that on the first day except that there was no explanation of the outline of the experiment, no consent form to complete, and no first questionnaire to complete. The session thus started with the training conversation.

E. Measurements

Subjects' impression of the robot was assessed using the scale proposed by Kanda and colleagues[10]. This scale consists of 25 semantic differential (SD) pairs that represent a variety of people's impressions of robots. For each pair of adjectives, subjects responded on a seven-point scale.

Subjects' trust in robots was measured with the scale proposed by Kidd[11]. The scale consists of six items. For each item, subjects responded on a seven-point Likert scale (1: strongly disagree; 7: strongly agree). The average of subject's responses to the six items was calculated and used as the trust score for that subject.

Finally, subjects' attitude towards robots was assessed with the Multi-dimension Robot Attitude Scale[12]. This scale consists of 49 question items assessing 12 sub-dimensions of general attitudes towards robots: familiarity, interest, negative attitude, self-efficacy, appearance, utility, cost, diversity, control, social support, operation, and environment fit. Subjects responded to each question item on a seven-point Likert scale (1: strongly disagree; 7: strongly agree). For each of the 12 sub-dimensions, the average of subject's responses to its member question items was calculated[12] and used as the score for that sub-dimension.

III. Results

A. Extraction of Dimensions of Impression

Even with the same set of measurement variables[10], it is still possible that the factor structure may be different across different sets of subjects and different experiments. Thus, a factor analysis was conducted on subjects' responses to the 25 SD scales of impressions to extract the dimensions of subjective impressions of the robot in the case of the present experiment. As each subject responded to 25 variables under two conditions (with and without speech errors), the number of samples was 88, which was more than three times the number of variables. All variables were checked with the normality of their distributions. No ceiling or floor effects were detected for any of the 25 variables.

A factor analysis employing the maximum likelihood method and promax rotation revealed a six-factor structure, with the criterion that eigenvalues be greater than zero. The

TABLE I

AN EXAMPLE OF THE CONVERSATIONAL SCRIPT USED IN THE EXPERIMENT (PART OF CONTENT 1). TEXT IS PRESENTED AS A ROMAN-CHARACTER EXPRESSION OF THE JAPANESE LANGUAGE. TEXT IN PARENTHESES IS ERRONEOUS SPEECH.

robot:	Kyou ha (*Kyo, kyou ha*) samui desune.
subject:	Samuine.
robot:	Samui hi ha atatakai (*akatatai, janakute, atatakai*) mono ga tabetai desune. Kinou no yoru ha nani wo tabemashitaka?
subject:	Nabe tabeta.
robot:	Nabe desuka. Iidesune, nabe. Fuyu no teiban desune. Anata ha naninabe ga sukidesuka?
subject:	Motsu-nabe.
robot:	Motsu-nabe desuka. Oishisou desune. Boku mo tabetemitakatta desu. Tokorode, nabe nitsuite hanashi wo shitemo iidesuka?
subject:	IIyo.

TABLE II

FACTOR LOADINGS OF SIX FACTORS OF IMPRESSION EVALUATIONS.

items	Factor 1 Familiarity	Factor 2 Amusement	Factor 3 Favorability	Factor 4 Agility	Factor 5 Sincerity	Factor 6 Presence
easy to get close to	0.930					
opening up	0.861					
approachable	0.767					
exciting		0.913				
amusing		0.726				
interesting		0.635				
cute		0.435	0.364			
bright						
smart			0.984			
amiable			0.659			
easy to understand			0.507			
good		0.365	0.433			
fond		0.363	0.418			
active			0.399			0.357
safe						
fast				0.895		
nimble				0.691		
tender				0.590		
vigor						
considerate					0.668	
friendly					0.505	
affable					0.370	
exotic						0.575
tough						0.530
cheerful						0.378

Note: Factor loadings whose absolute values are less than 0.35 are suppressed.

cumulative contribution of the six factors was 57.4%. Table II shows the factor matrix.

Items such as “easy to get close to”, “opening up”, and “approachable” showed high loadings for the first factor. This factor was therefore labeled “Familiarity”.

Items such as “exciting”, “amusing”, and “interesting” showed high loadings for the second factor. This factor was therefore labeled “Amusement”.

Items such as “cute”, “smart”, and “easy to understand” showed high loadings for the third factor. This factor was therefore labeled “Favorability”.

Items such as “fast”, “nimble”, and “tender” showed high loadings for the fourth factor. This factor was therefore labeled “Agility”.

Items such as “considerate”, “friendly”, and “affable” showed high loadings for the fifth factor. This factor was therefore labeled “Sincerity”.

Finally, items such as “exotic”, “tough”, and “cheerful” showed high loadings for the sixth factor. This factor was therefore labeled “Presence”.

These results were compared with a factor structure previously reported for the same set of SD scales[10]. The factor labeled “Familiarity” in the previous study appeared to be divided into two factors in the present study, namely “Familiarity” (factor 1) and “Favorability” (factor 3), suggesting that these two sets of variables represented independent dimensions of the impression among the subjects in the present study. Scores of the six factors are further analyzed as variables representing six dimensions of subject impressions of the robot.

B. Effect of Speech Errors on Impression

To compare subjects’ impressions of the robot between conditions with and without speech error, a series of four-way analyses of variance (ANOVAs) were conducted with the scores of the six factors extracted in section III-A and the trust score as seven characteristic variables, and the speech error, time of speech errors, order of contents, and sex as

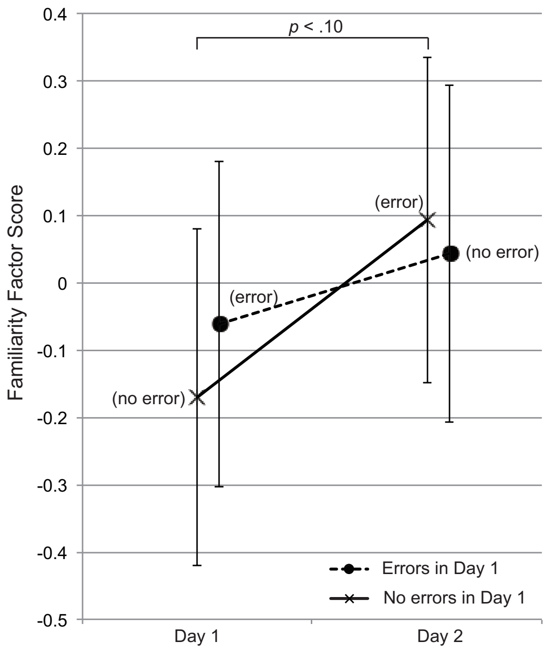

Fig. 1. Estimated means and standard errors of Familiarity factor scores by speech error conditions and time of speech error conditions.

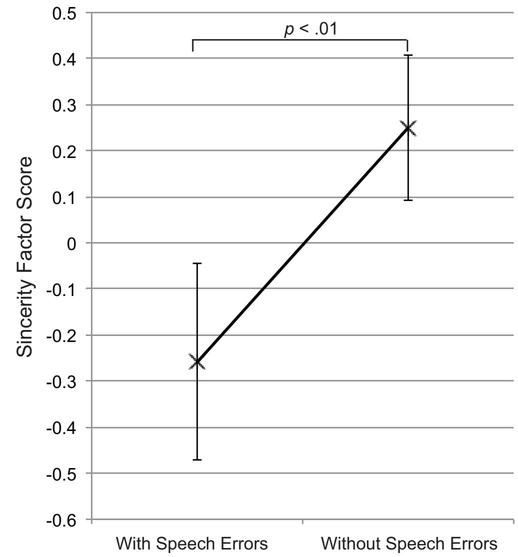

Fig. 2. Estimated means and standard errors of Sincerity factor scores by speech error conditions.

TABLE III

NUMBER OF SUBJECTS IN HIGH- AND LOW-ATTITUDE GROUPS. (N = 44)

	Number of subjects in groups	
Attitude dimension	Low-scoring-attitude group	High-scoring-attitude group
familiarity	19	25
interest	16	28
negative attitude	27	17
self-efficacy	21	23
appearance	19	25
utility	20	24
cost	25	19
diversity	22	22
control	18	26
social support	21	23
operation	26	18
environment fit	19	25

factors. Speech error was a within-subject variable, while the other three factors were between-subject variables.

For the Familiarity factor score, a moderately significant interaction between speech error and the time of speech error was found ($F = 3.06, df = 1, p < 0.10$), while none of the main effects were significant. Estimated means and standard errors of the Familiarity factor scores are shown in Fig. 1 by speech error conditions and time of speech error conditions.

Post-hoc analyses revealed that under the condition that the robot did not make speech errors on the first day and made errors on the second day, the Familiarity score was moderately higher when the robot made speech errors. In contrast, under the condition that the robot made speech errors on the first day and not on the second day, the Familiarity score was no different between speech error conditions. These results imply that people feel more familiarity with the robot that did not make any speech errors in their first contact, but then made speech errors afterwards. In contrast, if the robot made speech errors during the first interaction, the speech errors did not affect people's perception of familiarity significantly.

For the Sincerity factor score, a significant main effect of the speech error was found ($F = 10.86, df = 1, p < 0.01$). Estimated means and standard errors of the Sincerity factor scores are shown in Fig. 2 by speech error conditions. Post-hoc analyses revealed that the Sincerity score was significantly lower under the speech-error condition than under the without-speech-error condition.

No significant main effects or first-order interactions were found for the other four factor scores and the trust score.

C. Stratification by Attitudes toward Robots

Finally, to investigate whether the effects of speech errors on familiarity differ according to people's attitudes towards robots, subjects were divided into two groups holding different attitudes towards robots (e.g., attitudes relating to whether robots should be diversely applied or have varying functions), and the effects of the speech error on familiarity were compared across attitude groups.

For each of the 12 sub-dimensions of attitudes towards robots, subjects with attitude scores higher than the average were grouped into the high-attitude group, while those whose attitude scores were equal to or lower than the average were grouped into the low-attitude group. The numbers of subjects in these attitude groups are summarized in Table III.

For each of the high-attitude and low-attitude groups, an ANOVA was conducted with the Familiarity factor score as the characteristic variable and speech error, time of speech errors, order of contents, and sex as factors. Results were compared between high- and low-attitude groups.

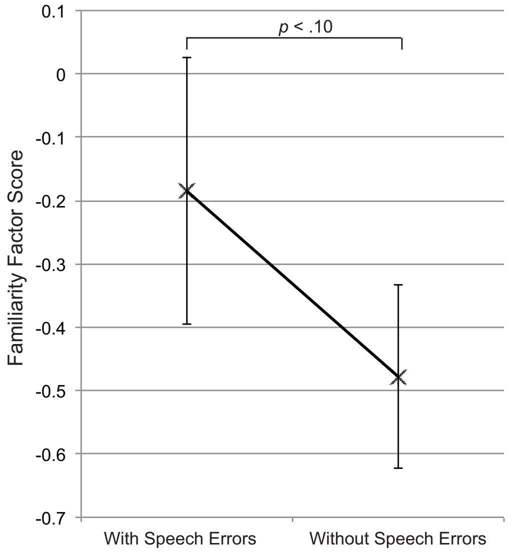

Fig. 3. Estimated means and standard errors of Familiarity factor scores of the low-diversity-attitude subjects by speech error conditions.

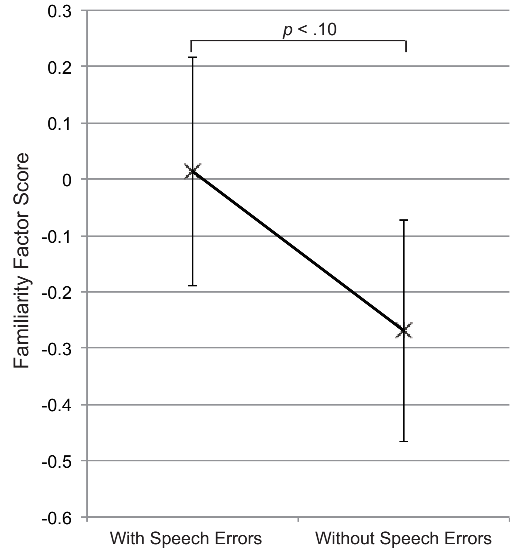

Fig. 4. Estimated means and standard errors of Familiarity factor scores of the low-operation-attitude subjects by speech error conditions.

In the group having low diversity attitude scores, a moderately significant main effect of speech error was found ($F = 3.776, df = 1, p < 0.10$). Estimated means and standard errors of the Familiarity factor scores of the low-diversity-attitude subjects are shown in Fig. 3 by speech error conditions. Post-hoc analysis revealed that the Familiarity factor score was moderately higher under the speech-error condition than under the non-speech-error condition. In contrast, in the group having high diversity attitude scores, neither significant main effects nor first-order interactions were found. This result implies that people holding the idea that robots should not have broad diversity would perceive robots with speech errors as more familiar.

In the group having low diversity attitude scores, the main effect of sex was also significant ($F = 4.732, df = 1, p < 0.05$). Male subjects had significantly higher Familiarity scores than female subjects.

In the group having low operation attitude scores, a moderately significant main effect of speech error was found ($F = 3.444, df = 1, p < 0.10$). Estimated means and standard errors of the Familiarity factor scores of the low-operation-attitude subjects are shown in Fig. 4 by speech error conditions. Post-hoc analysis revealed that the Familiarity factor score was higher under the speech-error condition than under the non-speech-error condition. By contrast, in the group having a high operation attitude score, there were no significant main effects or first-order interactions. This result implies that people who believe that the operations of robots should be limited and inflexible perceive robots with speech errors as being more familiar.

No significant main effects or first-order interactions were found in any of the other attitude subgroups.

IV. DISCUSSION

The present study investigated the effects of speech errors made by a robot on people's perception of the familiarity of the robot. It was found that speech error improved the familiarity score when the robot did not make any speech errors in the first contact with subjects and then made speech errors in the second contact on a separate day. However, speech errors lowered subjects' perception of the sincerity of the robot. These results imply that the robot should not make speech errors in the early stage of engagement with human users, while some speech errors after the users become accustomed to the robot might be effective in improving the familiarity of the robot.

It was also found that speech error improved familiarity for subjects with low-scoring attitudes in diversity and operation dimensions. These subjects are considered to have relatively fixed concepts about robots. A low score of the diversity attitude implies that the subject believes that there should not be a wide variety of robots. In the same way, a low score of the operation attitude implies that the subject believes that people should operate robots in rather fixed ways without flexibility. One possible explanation is that speech error might be effective in overcoming a fixed perception of robots, providing a novel surprise and thus improving the impression of robots. This implies that speech errors might be effective in improving familiarity, especially for people with rather fixed concepts of robots.

A limitation of the present study is that although there are several types of speech errors, differences across these types were not taken into consideration. Speech errors such as addition, drop, swap, and substitution are different in nature, and are also known to differ in frequency of occurrence. Thus, different types of speech errors might have different effects on familiarity and other impressions of robots. Investigating the difference across types of speech errors should be pursued in future studies.

All subjects in this experiment were graduate and undergraduate students who have a background in science and technology. Subjects having a broader range of ages, more diverse backgrounds, and/or different cultures should be included in further work to further generalize the implications of the present study.

Finally, in the present study, subjects interacted with robots only for two days and only for a period as short as minutes on each day. In reality, it is likely that users will engage with robots for a much longer period after first contact. Each interaction can also be much longer than several minutes. Long-term studies observing changes in the users' engagement with and impressions of robots over a longer period would be worthwhile.

REFERENCES

[1] A. Kim, J. Jung, K. Lee, and J. Han, "The effects of familiarity and robot gesture on user acceptance of information," in *Proc. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI2013)*, Tokyo, Japan, Mar. 2013, pp. 159–160.

[2] M. Okada, "Human-like compliant motion and its mechanism," *J. Robotics Society Japan*, vol. 17, no. 6, pp. 782–785, 1999.

[3] A. Sakamoto, Y. Hayashi, and H. Ogawa, "Expression of human-likeness with communication robot PaPeRo," in *Proc. Human-Agent Interaction Symposium 2011*, Kyoto, Japan, Dec. 2011.

[4] J. Austin, Ed., *How to do things with words*. Oxford: Clarendon Press, 1962.

[5] Y. Terao, Ed., *Iimachigai wa doushite okoru? (How speech errors occur?)*. Tokyo: Iwanami Shoten Publishers, 2002.

[6] E. Saito, "Kikimachigai no kigeki (Comedy of hearing wrong)," *Izumia Liberalia : Annals of Graduate School of Humanities Meiji University*, vol. 1, pp. 92–96, 2009.

[7] T. Nomura, T. Kanda, and T. Suzuki, "Experimental investigation into influence of negative attitudes toward robots on human-robot interaction," *AI & Society*, vol. 20, no. 2, pp. 138–150, 2006.

[8] NAO robot: intelligent and friendly companion — Aldebaran. [Online]. Available: https://www.aldebaran.com/en/humanoid-robot/nao-robot

[9] SDK, simple software for developing your NAO — Aldebaran. [Online]. Available: https://www.aldebaran.com/en/robotics-solutions/robot-software/development

[10] T. Kanda, H. Ishiguro, and T. Ishida, "Psychological evaluations on interactions between people and robot," *J. Robotics Society Japan*, vol. 19, no. 3, pp. 362–371, 2001.

[11] C. D. Kidd, "Sociable robots: The role of presence and task in human-robot interaction," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2003. [Online]. Available: http://web.media.mit.edu/˜coryk/papers/Kidd_MS_thesis.pdf

[12] A. Fujita, D. Suzuki, and H. Umemuro, "Cross-regional comparative study of dimensions of people's attitudes toward robots," in *Proc. 22nd IEEE International Symposium on Robot and Human Interactive Commnication (RO-MAN2013)*, Gyeongju, South Korea, Aug. 2013, p. 18.