Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Production Experiments on Two Cases of Tonal Neutralization in Taiwan Southern Min

Mao-Hsu Chen

University of Pennsylvania, USA
chenmao@sas.upenn.edu

Abstract

This study explored two cases of tonal neutralization in Taiwan Southern Min: 1) context tones 55 and 24 which are both realized as sandhi tone 33 on surface when occurring in context positions, and 2) context tone 21 and context checked tone 21 with a glottal stop coda are said to be realized the same in context positions as a high falling sandhi tone 51. Speakers of two age groups, younger and older, were recruited for the production experiments to examine whether the neutralization is complete. Comparison between the four quartile and overall mean f0 values showed an age-based acoustic variation in the first case, where context tones 55 produced by the older speakers were significantly higher in pitch than context tones 24 throughout the entire contour, which was absent from the younger speakers data. The second case, however, revealed a case of complete neutralization in terms of both the f0 contours, where no significant difference in the quartile and overall mean f0 values was found, and the durations of the target syllable as reported by linear mixed effects modeling.

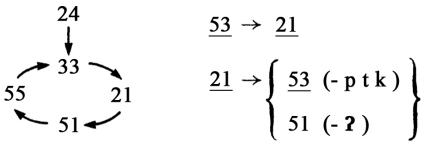
Index Terms: tontal neutralization, tone sandhi, incomplete neutralization, Taiwan Southern Min

1. Introduction

Phonological neutralizations have been proposed to be incomplete, posing a threat to formal phonological theory as to what phonological and phonetic information account for the mental representation of a lexical item. A strong claim came from Port & Crawford's [1] study on the acoustic contrast of German voiced and voiceless stops at syllable-final position. For example, a devoiced obstruent in *Bund* 'group, association' is said to become indistinguishable from its underlyingly voiceless counterpart, say the final stop consonant in *bunt* 'colorful'. Their discriminant analysis and identification task, however, showed significant differences in both production and perception, supporting their idea that the final devoicing in German must be represented not only with a categorical phonological rule that signals the discrete change in the voicing feature, but also with a phonetic implementation rule that justifies continuously variable values of phonetic features. Other effects of incomplete neutralization have been found in final devoicing in Catalan [2], Polish [3], Russian [4], Turkish [5], and Dutch [6].

While most of the early neutralization studies have been focused exclusively on the encoding of segmental information, the interest seems to be extended into autosegmental information, allotones or tone sandhi phenomena, in particular. Peng [7] investigated the Mandarin third tone sandhi in which tone 3 (214) becomes tone 2 (35) when followed by another tone 3. Note that the numerical values in parentheses represent pitch height on a five-point scale introduced by Chao [8], where 5 indicates the highest pitch and 1 the lowest. The production experiment found marginally significant effect of tone type (underlying tone 2 vs. the sandhi tone) on the mean f0 values; nevertheless, native speakers in the identification experiment failed to perceive the small pitch difference found in the production experiment, suggesting that the sandhi tone is not completely neutralized with the underlying tone 2 acoustically but is perceptually neutralized to some degree such that it is indistinguishable from tone 2.

Taiwan Southern Min (TSM), also known as Taiwanese, Xiamen, Amoy, and Hokkien, is noted for its complex tone sandhi (TS) system. Every lexical word in TSM has a base (also called citation/juncture) tone and a sandhi tone. The tone sandhi referring to the alternations between these two tones depends solely on the position of the word within a prosodic constituent, a tone group (TG). A word is realized with the base tone when occurring in the juncture position. i.e., the right edge of a TG; when occurring elsewhere, it is realized with the sandhi tone. TSM has seven contrastive tones, including five free tones and two checked tones with CV[p, t, k, h[1]] syllable structure that are usually characterized by shorter duration and glottalized voice quality [9, 10]. The tone sandhi systems of the two group are illustrated in Figures 1(a) and 1(b), the former often referred to as the Tone Circle as suggested by the circular movement. The checked tones are marked with underlines, and the direction of the arrow shows the selection of the surface sandhi form. For instance, the sandhi form of a lexical high level tone 55 is a mid level tone 33, the sandhi form of a lexical mid level tone 33 is a low falling tone 21, and so on.



(a) Free syllables (b) Checked syllables

Figure 1: *The tone sandhi of free syllables (Tone Circle) and that of checked syllables.*

As can be inferred from Figure 1, there exist two possible cases of tonal neutralization in the convoluted tone sandhi system in TSM: 1) both context tones 55 and 24 are realized as sandhi tone 33 on surface when occurring in context positions, and 2) context tone 21 and context checked tone 21 with a glottal stop coda are realized as a high falling sandhi tone 51 in context positions. Tsay et al. [11] examined the neutralization of context 55 and context 24 within disyllabic compounds and

[1]The letter *h* is traditionally used for a glottal stop /ʔ/ in Chinese linguistics.

 doi: 10.21437/SpeechProsody.2016-24

showed that the alternation between juncture and context tones was categorical, as indicated by the neutralization of the f0's of juncture and context forms, confirming the categoricality hypothesis. Myers and Tsay [12] employed minimal pairs of sentences to examine whether neutralization occurred in two conditions: a) across-positionally (juncture vs. context), in which the target words occupied different positions in a TG, one in the juncture position (realized with its base tone 33) while the other in the context position (realized as the sandhi tone 33 derived from base tone 55), and b) within-position (context vs. context), where the pairs of sentences had the same TG formation while the target words had different underlying tones (55 or 24) but the same surface tones 33. Discourse contexts (listener-absent or -present) were also taken into consideration. No significant effect of discourse context on duration and f0 was found, and the difference in overall f0 between context 55 and context 24 was not only insignificant but also extremely small, a mere 2.3 Hz. The duration with context 24 within context position is significantly longer, 8 ms on average, than context 55. No overall difference in slope between these two tones in context position was found, contrary to the claim about incomplete neutralization that the processes giving rise to incomplete neutralization involved the temporal adjustment of gestures and their relations [13, 14, 15].

Production experiments were conducted to investigate the two cases of tonal neutralization in TSM with respect to 1) whether the two mid level sandhi tones (33), one derived from the high level tone (55) and the other from the low rising tone (24) are completely neutralized in the same context position, and 2) whether there is an acoutical difference between the two high falling snadhi tones (51), one originated from the unchecked low falling tone (21) and the other from the checked version of low falling tone (21). For expository convenience, we refer to the first tonal neutralization case as sandhi-33 and the second one as sandhi-51, as inspired by their corresponding surface realizations.

2. Method

2.1. Subjects

13 native speakers of TSM, 6 males and 7 females, were recruited for both production experiments. They were divided into two age groups, seven speakers, 4 males and 3 females, in the younger group (ranging from 25 to 35 years old with an average of 28.43) and six speakers, 2 males and 4 females, in the older group (ranging from 52 to 66 years old with an average of 58.83). All the older speakers and two of the younger speakers were recorded in Taiwan, while the others were recorded at the University of Pennsylvania. Those sampled from Pennsylvania were Taiwanese graduate students studying in Philadelphia and they have stayed in the US for no more than three years. All subjects were able to speak and read Mandarin.

2.2. Instruments

Recordings were made either with Audacity on computers with Mac or Windows OS and the Shure WH30 condenser headset microphone, or with a recording application using 44.1 kHz/16-bit sampling rate on a smartphone with the built-in microphone.

2.3. Materials

2.3.1. sandhi-33

The production experiment on the sandhi-33 case included 10 target minimal pairs of monosyllablic words, all of which were embedded in minimal pairs of sentences, as exemplified in (1) and (2), where the target words are in bold and are marked with their base tones, 24 and 55, but they are both realized as 33 on the surface, i. e., the sandhi tone. Each of the stimuli sentences had a length of four to ten syllables with an average length of 6.23 syllables. All the pairs of target words were at sentence-medial position. A randomized reading list including 300 sentences ((10 minimal pairs $\times$ 2 words + 10 fillers) $\times$ 10 repetitions) was generated for each subject.

(1) beh kio i **poe24** kau tang-si
want ask him accompany until when
'How long do you still want him to accompany you?'

(2) beh kio i **poe55** kau tang-si
want ask him fly until when
'How long do you still want him to travel by flight?'

2.3.2. sandhi-51

Only five target minimal pairs of monosyllablic words were used in the production experiment on the sandhi-51 case due to the difficulty in finding minimal pairs and the low proportion of checked tones in TSM. The target words were embedded in minimal pairs of sentences, similar to (1) and (2), with an average length of 5.2 syllables (ranging from four to seven syllables), and were located at neither sentence-initial nor -final position with one exception at sentence-initially. To counterbalance the stimuli, a randomized reading list of 100 sentences ((5 minimal pairs $\times$ 2 words + 10 fillers) $\times$ 5 repetitions) was generated for each subject.

2.4. Procedure

During the recording, one experimenter was either on-site or remotely monitoring the experiment to provide detailed instructions and ensure that the participants understood and performed the task as expected. The randomized stimuli were presented one by one with Chinese characters on the computer screen to elicit subjects production. Subjects were asked to first read the number of the order for each token followed by the stimuli sentence as naturally as possible and they were allowed to rest at any time. For the sandhi-33 case, the recording lasted for about 45 minutes with 300 tokens recorded from each speaker, 200 of which were the intended target used in further analyses. The experiment for the sandhi-51 case took about 15 minutes for each participant, and 50 out of the 100 recorded tokens were involved in the analyses of f0 contour and duration of the target syllable.

2.5. Data analysis

The nucleus of each target syllable was hand-labeled in Praat [16]. F0 values at 10 and 12 equally-spaced points in the rhyme of each target syllable were extracted with VoiceSauce [17] and converted into semitones using 100Hz as the reference. Each f0 contour was further divided into four regions, and the mean f0 of each quartile as well as the overall mean f0 were calculated. Each f0 contour was therefore represented by five mean values for regression analysis. The data of the two age groups were analyzed separately in order to see the generational difference.

In addition, the boundaries of the sentences and the tar-

get syllables were manually marked to calculate the normalized duration of each target syllable as a percentage of the sentence duration for the sandhi-51 case, since the fundamental difference in syllable durations between checked and unchecked tones could potentially be a source of incomplete neutralization.

As some tokens were pronounced differently from what were expected, they were excluded from the analyses. For the sandhi-33 case, a total of 2,389 f0 contours were analyzed, including 1,258 tokens from the younger group and 1,131 tokens from the older group, with the repeated-measures linear mixed effects model and the Likelihood Ratio Test in R [18] using the *lme4* package [19]. Base tone and gender were the fixed effects while subject and token were the random effects. A random slope model was adopted to account for inter-subject and inter-token variations. Interactions between base tone and gender were also included. The resulting full and reduced models can be summarized as in (3), (4) and (5), respectively. For the sandhi-51 case, 649 f0 contours were extracted, including 350 tokens from the younger group and 299 tokens from the older age group, and the same set of repeated-measures linear mixed effects models and the Likelihood Ratio Test were used for the analysis of f0 contours.

(3) Full: mean $\sim$ base tone * gender + (1 + base tone | subject) + (1 + base tone | token)

(4) Reduced1: mean $\sim$ base tone + gender + (1 + base tone | subject) + (1 + base tone | token)

(5) Reduced2: mean $\sim$ gender + (1 + base tone | subject) + (1 + base tone | token)

Utterances with pause or hesitation in between were further excluded for the analysis of duration. In total 598 normalized durations were calculated, including 320 tokens from the younger group and 278 tokens from the older group. The data of the two age groups were analyzed together here as no obvious generational difference in duration was perceived from Figure 4. Another set of repeated-measures linear mixed effects models, where *age* was regarded as a fixed effect in addition to base tone and gender in the full model as exemplified in (6), were applied to examine the durational difference. For each iteration of the Likelihood Ratio Test, only one of the fixed effects or random effects was dropped to produce a further reduced model, and the final best model for the duration data is shown in (7).

(6) Full: dur $\sim$ base tone * gender * age + (1 | subject) + (1 | token)

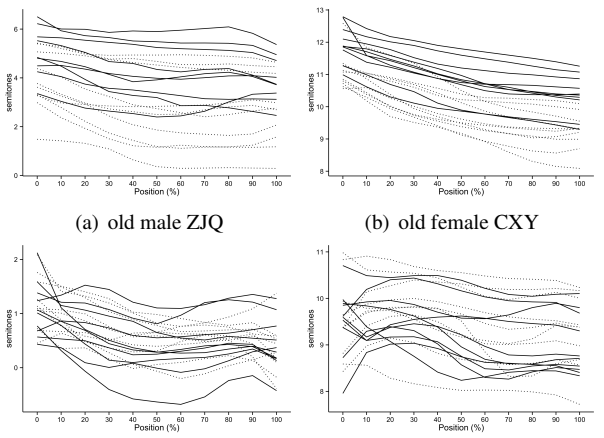
(7) Reduced: dur $\sim$ 1 + (1 | subject) + (1 | token)

2.6. Results

2.6.1. sandhi-33

Figure 2 illustrated the difference in the realizations of the sandhi tone 33 between different age groups. In Figures 2(a) and 2(b), most of the solid lines were above the dotted lines, suggesting that the older speakers tended to pronounce the target words of base tone 55 with a higher pitch than those of base tone 24. As compared with the lower two graphs, one could observe that the sandhi tones 33 derived from base tones 55 and 24 produced by the older speakers were more distinct from each other than those produced by the younger speakers, where more overlaps were found.

The result of the Likelihood Ratio Test comparing reduced1 with full model indicated that the interaction term between base



(a) old male ZJQ (b) old female CXY

(c) young male TSY (d) young female CSH

Figure 2: *The f0 contours of the token "poe" produced by one male and one female speakers in each age group. The solid lines indicate the f0 contours for tokens with tone 55 as the base tone whereas the dotted lines refer to the f0 contours of tokens with base 24 tone.*

tone and gender for the younger age group is not significant, and thus there is no interdependence between the two fixed effects. For the comparison between reduced1 and reduced2 models, Table 1 suggested that in all cases of the mean f0 of each target syllable among the younger speakers, the underlying tone shows no significant effect on the surface f0 values.

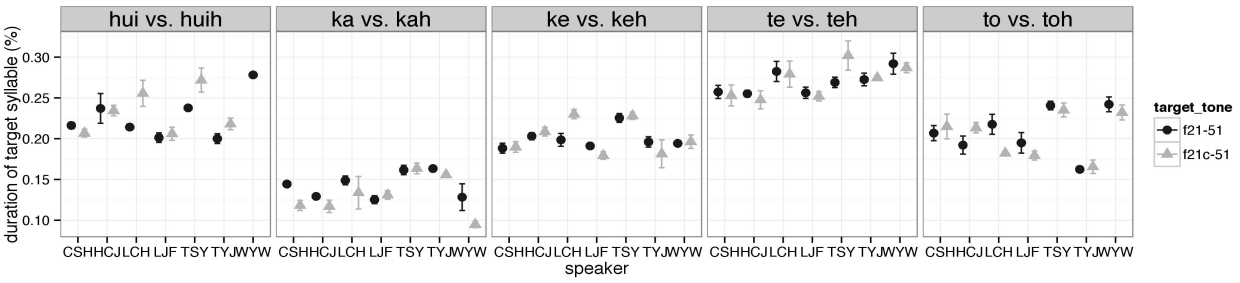
Table 1: *The statistical result of the effect of the base tone on the mean f0 values for the younger age group, with the reduced2 model compared with the reduced1 model.*

Result	Q1	Q2	Q3	Q4	All
p-value	0.130	0.111	0.068.	0.066.	0.069.
$\chi^2(1)$	2.297	2.545	3.340	3.388	3.310

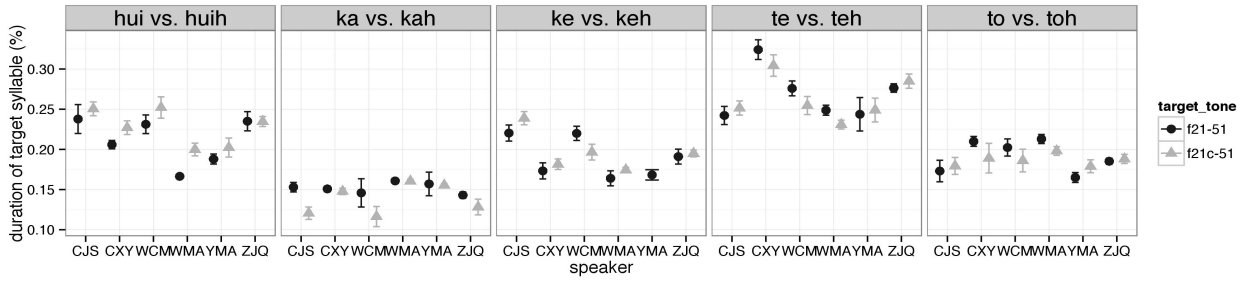
Table 2 showed the result of the Likelihood Ratio Test when comparing the two reduced models in (4) and (5), where the interaction term between base tone and gender for the older age group is significant, suggesting that base tone has a significant effect on all the mean f0 values that were somehow modulated through gender, i. e., base tone and gender were significantly inter-dependent on each other. The by-subject and by-token coefficients for the effect of base tone were different for each subject and token, as estimated in our random slopes model. However, the values were always positive and that many of them were quite similar to each other, revealing that despite individual variation, there was also consistency in how base tone affected all the mean f0 values; that is, for all the older speakers, all the five mean f0 values went up when the base tone of the token was 55, but for some people they went up slightly more so than for others.

2.6.2. sandhi-51

Figure 3, on the other hand, showed no clear difference in the realizations of the sandhi 51 tones between different age groups, where the solid lines overlapped with the dotted lines, which is compatible with the results of the Likelihood Ratio Test, where

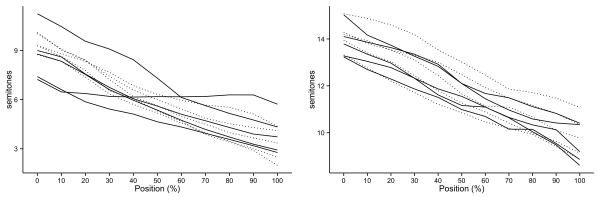


(a) younger speakers



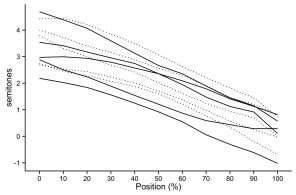
(b) older speakers

Figure 4: *Normalized durations of the target syllable as a proportion to the sentence duration. The point indicates the mean duration while the error bars represent ± standard deviation from the mean. The darker marks denote sandhi 51 tones derived from unchecked 21 while the lighter ones are for sandhi 51 with underlying checked 21 tones.*

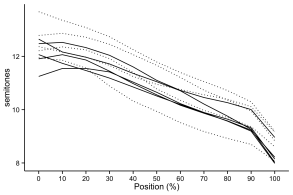


(a) old male ZJQ

(b) old female CXY



(c) young male TSY



(d) young female CSH

Figure 3: *The f0 contours of the minimal pair "to" and "toh" produced by one male and one female speakers in each age group. The solid lines indicate the f0 contours for tokens with checked tone 21 as the base tone whereas the dotted lines refer to the f0 contours of tokens with base unchecked 21 tone.*

no significant effect on the surface f0 values was found throughout the entire f0 contour for both age groups (all p-values $>$ 0.1626).

Figure 4 showed the difference in durations of the two sandhi 51 tones in each minimal pair for each subject in the younger and older groups, where no obvious generational pattern but individual difference and sensitivity to stimuli can be observed. The Likelihood Ratio Tests indicated that only the two random effects, subject ($\chi^2(1)$=64.938, p = 7.731e-16) and token ($\chi^2(1)$=731.63, p $<$ 2.2e-16) are significant.

Table 2: *The statistical result of the interaction between base tone and gender for the older age group, with the reduced1 model compared with the full model.*

Result	Q1	Q2	Q3	Q4	All
p-value	0.026*	0.006**	0.014*	0.018*	0.009**
$\chi^2(1)$	4.9603	7.405	6.035	5.574	6.829

3. Discussion

For the sandhi-33 case, distinct production patterns are observed for different age groups in terms of the effect of the base tone on the surface f0 values. Younger speakers do not maintain a pitch distinction between the two sandhi tones 33, one derived from base tone 55 and the other from base tone 24. Older speakers, on the other hand, preserve the difference in the underlying forms and consistently produce the sandhi tone 33 with 55 as the base tone with higher pitch than the one derived from base tone 24. As incomplete neutralization is closely related to the phenomenon of near mergers, the age-based acoustic variation could also signal an ongoing change towards complete merger, which is worthy of a longitudinal study.

The sandhi-51 case, however, demonstrates a complete tonal neutralization in terms of the f0 contours and the durations of the target syllables. Two possible interpretations of this complete merger could be suggested. First, the neutralization of checked and unchecked 21 tones in the context environment is in essence fully phonological, and thus only by-subject and by-token differences are available. The second possibility would be that the neutralization was phonetically incomplete but that the

process of a change towards complete merger has completed. Speakers of older generation would have to be recruited to support this idea.

The production experiments on the two cases of tonal neutralization in Taiwan Southern Min also showed that the neutralizations of different tone sandhi rules in a language could differ in terms of the completion of merger.

4. References

[1] R. F. Port and P. Crawford, "Incomplete neutralization and pragmatics in german," *Journal of Phonetics*, vol. 17, no. 4, pp. 257–282, 1989.

[2] J. Charles-Luce and D. A. Dinnsen, "A reanalysis of catalan devoicing," *Journal of Phonetics*, vol. 15, pp. 187–190, 1987.

[3] L. M. Slowiaczek and D. A. Dinnsen, "On the neutralizing status of polish word final devoicing," *Journal of Phonetics*, vol. 13, no. 3, pp. 325–341, 1985.

[4] A. J. Olga Dmitrieva and J. Sereno, "Phonological neutralization by native and non-native speakers: The case of russian final devoicing," *Journal of Phonetics*, vol. 38, pp. 483–492, 2010.

[5] C. Rudin, "Phonetic evidence for a phonological rule: G-deletion in turkish," *Research in Phonetics*, vol. 1, pp. 217–232, 1980.

[6] N. Warner, A. Jongman, J. Sereno, and R. Kemps, "Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from dutch," *Journal of phonetics*, vol. 32, no. 2, pp. 251–276, 2004.

[7] S.-H. Peng, "Lexical versus phonological representations of Mandarin Sandhi tones," *Papers in laboratory phonology V: Acquisition and the lexicon*, pp. 152–167, 2000.

[8] Y. R. Chao, *Mandarin primer: An intensive course in spoken Chinese*. Harvard University Press, 1948.

[9] R. L. Cheng, "Tone sandhi in Taiwanese," *Linguistics*, vol. 6, no. 41, pp. 19–42, 1968.

[10] ——, "Some notes on tone sandhi in Taiwanese," *Linguistics*, vol. 11, no. 100, pp. 5–25, 1973.

[11] J. Tsay, J. Charles-Luce, and Y.-S. Guo, "The syntax-phonology interface in taiwanese: acoustic evidence," in *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 3, 1999, pp. 2407–2410.

[12] J. Myers and J. Tsay, "Neutralization in taiwan southern min tone sandhi," *Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday*, pp. 47–78, 2008.

[13] E. Zsiga, "Features, gestures, and the temporal aspects of phonological organization," Ph.D. dissertation, 1993.

[14] R. F. Port, "The discreteness of phonetic elements and formal linguistics: response to a. manaster ramer," *Journal of Phonetics*, vol. 24, no. 4, pp. 491–511, 1996.

[15] R. F. Port and A. P. Leary, "Against formal phonology," *Language*, pp. 927–964, 2005.

[16] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*, version 5.4.06, retrieved 21 February 2015. [Online]. Available: http://www.praat.org/

[17] P. K. Yen-Liang Shue and C. Vicenik, "Voicesauce: A program for voice analysis," *Journal of the Acoustical Society of America*, vol. 126, p. 2221, 2009. [Online]. Available: http://www.ee.ucla.edu/spapl/voicesauce

[18] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[19] D. Bates, M. Maechler, B. Bolker, and S. Walker, *lme4: Linear mixed-effects models using Eigen and S4*, r package version 1.1-7. [Online]. Available: https://github.com/lme4/lme4/ http://lme4.r-forge.r-project.org/

[20] J. B. Grier, "Nonparametric indexes for sensitivity and bias: computing formulas." *Psychological bulletin*, vol. 75, no. 6, p. 424, 1971.