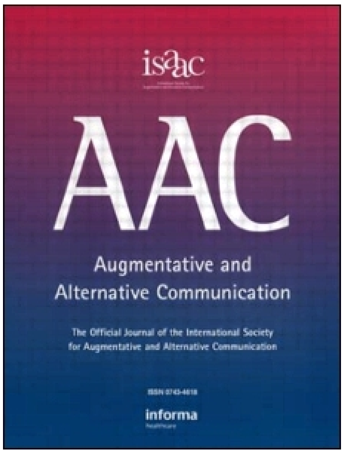

Augmentative and Alternative Communication
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713692248

Computerized speech recognition: influence of intelligibility and perceptual consistency on recognition accuracy

Nancy Thomas-Stonell [a]; Ava-Lee Kotler [b]; Herbert A. Leeper [c]; Philip C. Doyle [c]
[a] Graduate Department of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada. [b] Bloorview MacMillan Centre, Toronto, Ontario, Canada. [c] Communication Disorders Program, University of Western Ontario, London, Ontario, Canada.

Online Publication Date: 01 March 1998

To cite this Article Thomas-Stonell, Nancy, Kotler, Ava-Lee, Leeper, Herbert A. and Doyle, Philip C.(1998)'Computerized speech recognition: influence of intelligibility and perceptual consistency on recognition accuracy',Augmentative and Alternative Communication,14:1,51 — 56

To link to this Article: DOI: 10.1080/07434619812331278196

URL: http://dx.doi.org/10.1080/07434619812331278196

0743-4618/98/1401-0051 $3.00/0; Volume 14, March 1998
AAC Augmentative and Alternative Communication

RESEARCH NOTE

Computerized Speech Recognition: Influence of Intelligibility and Perceptual Consistency on Recognition Accuracy

Nancy Thomas-Stonell, Ava-Lee Kotler, Herbert A. Leeper, and Philip C. Doyle
Graduate Department of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada (N.T.), Bloorview MacMillan Centre, Toronto, Ontario, Canada (A.K.), Communication Disorders Program, University of Western Ontario, London, Ontario, Canada (H.A.L., P.C.D.)

The effects of intelligibility and consistency on the recognition accuracy of a speaker-adaptable speech recognition system (IBM VoiceType Version 1.0) were evaluated. Six participants who had dysarthria of speech across three severity levels (i.e., mild, moderate, severe) and six age- and gender-matched peers without speech impairments participated in the study. Productions of sentences were evaluated across five assessment sessions. Recognition accuracy was significantly higher for the speakers in the control group than for the speakers with dysarthria across severity levels. High levels of intelligibility correlated significantly with high recognition accuracy scores. Perceptual rankings of speech consistency did not correlate significantly with recognition accuracy scores. Results suggest that for speaker-adaptable systems, the more intelligible a speaker, the greater his or her success with the voice recognition system. Results also suggest that perceived inconsistencies in the speech productions of dysarthric speakers may not limit their use of a speaker-adaptable speech recognition system.

KEY WORDS: computer recognition, dysarthria, speech impairment, speech recognition, voice recognition

Speech recognition is a practical computer access mode for people with motor limitations without speech involvement (Goodenough-Trepagnier & Rosen, 1991). Many people with disabilities affecting the upper limbs also have speech impairments. Some individuals with motor and speech impairments use their partially intelligible speech to communicate and to access computers (Sy & Deller, 1989). There is, however, little documented research concerning the use of speech recognition technology by individuals with speech impairments (Ferrier, Jarrell, Carpenter, & Shane, 1992). The available research indicates that dysarthric speakers are less well recognized by speech recognition systems than speakers who do not have speech impairments or speakers with hearing impairment (Carlson & Bernstein, 1987; Coleman & Meyers, 1991).

There are two main types of speech recognition systems: speaker-dependent and speaker-adaptable systems. Speaker-dependent speech recognition systems require the user to pretrain all words to be used with the system. Speaker-adaptable systems combine features of both systems (Ferrier et al., 1992). For speaker-adaptable systems, key words, used to control the system, are pretrained by the speaker before use. These systems also contain a set of pretrained word templates. Speakers' productions are matched against the word templates stored in memory. As the system is used, it adapts to individual speech characteristics by adding the individual's own word templates to its pretrained templates. This allows the system to learn the acoustic characteristics of individual speakers' speech (Stern & Lasry, 1987).

A recent literature review revealed two studies that evaluated the recognition accuracy of a speaker-adaptable speech recognition system (i.e., DragonDictate). Ferrier et al. (1992) demonstrated that recognition accuracy for a speaker with mild dysarthria approached the recognition accuracy achieved by those with normal speech. Recognition accuracy scores for the speaker with mild dysarthria were 10% to 15% lower and showed more variability across sessions. In a second study, the correlations between intelligibility measures and recognition accuracy were

TABLE 1: **Participant Profiles—Dysarthria Group Participants**

Participant Code	*Severity Subgroup**	*Gender*	*Age*	*Diagnosis*	*CAIDS†*	
					%W‡	*%S#*
P1	Mild	Female	19	TBI§	76	99
P2	Mild	Male	18	CP**	74	95
P3	Moderate	Male	23	TBI	42	88
P4	Moderate	Female	55	CP	40	85
P5	Severe	Female	34	TBI	38	90
P6	Severe	Male	15	CP	28	55

*Severity subgroup refers to ratings of mild, moderate, or severe dysarthria of speech based on CAIDS word task intelligibility score. †CAIDS = Computerized Assessment of Intelligibility of Dysarthric Speech (Yorkston et al., 1984). ‡%W = Word task intelligibility score. #%S = sentence task intelligibility score. §TBI = traumatic brain injury. **CP = cerebral palsy.

strong for the 10 participants with spastic cerebral palsy and varying levels of dysarthric speech (Ferrier, Shane, Ballard, Carpenter, & Benoit, 1995). All but three participants reached the 80% recognition accuracy criterion for dictation of the same passage within eight dictations. For speaker-dependent systems, consistency of sound production is paramount (Schmitt & Tobias, 1986). For speaker-adaptable systems, speech intelligibility may be at least as important as perceptual judgments of consistency of speech production in achieving good recognition accuracy scores.

A controlled evaluation of International Business Machines' (IBM) VoiceType (Version 1.0), a speaker-adaptable speech recognition system, was conducted. IBM's VoiceType was developed in conjunction with Dragon Systems Inc., which makes a similar speaker-adaptable speech recognition system called "DragonDictate" (IBM, 1991).[1] Recognition accuracy of VoiceType was examined with six individuals who had dysarthria of speech across three severity levels (i.e., mild, moderate, severe) and six age- and gender-matched peers. This study evaluated:

1. *Recognition accuracy:* Is recognition accuracy for sentences significantly poorer for people with varying degrees of dysarthria as compared to those who have no speech impairments?
2. *Intelligibility/consistency:* Is recognition accuracy for people with dysarthria correlated with speech intelligibility (Computerized Assessment of Intelligibility of Dysarthric Speech [CAIDS]) and/or perceptual measures of speech consistency?

METHOD

Participants

Twelve individuals were selected for participation in the study, ranging in age from 15 to 55 years (M = 27 years). This age range was selected so that participants' speech would be stable (i.e., past the developmental stage of speech acquisition). Selection criteria included hearing within normal limits and the ability to follow directions. Six individuals (three men, three women) had dysarthria of speech (developmental or acquired) and were selected from the client population of Bloorview MacMillan Centre (dysarthria group). Those participants who had acquired dysarthria secondary to traumatic brain injuries were at least 4 years post trauma. This also ensured stable speech patterns. See Table 1 for demographic data.

Selection for the dysarthria group ensured two participants (one male, one female) with mild, moderate, and severe intelligibility deficits. Speech intelligibility was measured by transcription using CAIDS (Yorkston, Beukelman, & Traynor, 1984). A mild intelligibility deficit was defined as intelligibility scores greater than 70% but less than or equal to 90% on the single-word task of the CAIDS. A moderate intelligibility deficit was defined as scores greater than 40% and less than or equal to 70%. A severe intelligibility deficit was defined as scores greater than 10% but less than or equal to 40%. Speakers whose intelligibility was less than or equal to 10% or greater than 90% were not included in the study in order to reduce floor and ceiling effects. Six individuals matched by age and gender, without articulation or voice deficits (control group), also participated in the project.

Design

All participants were seen for five assessment sessions within a 4-week period using the VoiceType system. During each session, participants read the words from the single-word intelligibility test proposed by

[1]Although VoiceType (Version 1.0) is no longer being produced by IBM, it should be noted that VoiceType was developed in conjunction with Dragon Systems Inc., which is still producing similar speech recognition software (i.e., DragonDictate).

Kent, Weismer, Kent, and Rosenbek (Kent words; 1989) and the first three sentences of the Rainbow Passage (Rainbow sentences; Fairbanks, 1960). The order of the Kent words was varied for each session. Productions were recorded simultaneously on the speech recognition system using a headset microphone and on a digital audiotape (DAT) recorder (Sony DTC-750) using a second microphone routed to the DAT recorder through a preamplifier. None of VoiceType's parameters (e.g., pause time) were adjusted.

Participants spoke each word in isolation at their habitual pitch and loudness levels. If a word was correctly recognized, the participants continued with the next word. When a word was not recognized correctly, the appropriate corrections were entered into VoiceType to ensure maximal adaptation of the system to the participant's speech patterns. If VoiceType did not respond to a spoken word, the word was repeated into VoiceType and recorded again on the DAT recorder. Throughout all sessions, VoiceType's choices in response to participants' productions were noted on a data sheet.

The recordings of the Kent words from the dysarthria group were also evaluated by 10 listeners. The listeners were staff or student volunteers from a university communication disorders program who had passed a pure-tone audiometric screening at 20 dB HL for the octave frequencies 500, 1000, and 2000 Hz. Listeners were requested to identify the word they perceived using a modified closed-set response (e.g., target and three foils) paradigm (see the Appendix for examples). In addition, there was an opportunity for the listener to write in the perceived word if it was not one of the foil items presented. The results of these perceptual evaluations were used to derive the intelligibility and consistency rankings.

A clinical measure of speech consistency was calculated, based on listeners' perceptions of clients' speech variability across the five training sessions. Although listeners' perceptions of speaker's productions may not correspond with acoustic correlates of consistency, the majority of clinicians do not have either the time to obtain acoustic measures of speech consistency or access to this type of equipment when evaluating clients for the potential use of speech recognition technology (Kent, 1996). Therefore, clinicians use perceptually based judgments of the consistency of client's speech productions to make decisions about the feasibility of speech recognition devices for clients.

Data Analyses

Goal #1: Recognition Accuracy

The number of words from the Rainbow sentences that were correctly recognized by VoiceType on the fifth training session was obtained for both the dysarthria and control groups. Percent accuracy scores (recognition accuracy) were calculated (i.e., number of words correctly recognized divided by the total number of words). The fifth session was selected because IBM indicated that five repetitions would allow for maximal adjustment of the VoiceType system (J. Roberts, personal communication, September 1991). An analysis of variance (ANOVA) was completed to determine if there were differences in the average recognition accuracy scores between the dysarthria and control groups. The ANOVA also examined differences between the control group and the subgroups (mild, moderate, and severe) of the dysarthria group. The ANOVA provided a profile analysis of the amount of change in recognition accuracy scores across sessions for both groups.

Goal #2: Intelligibility/Consistency

Participants from the dysarthria group were rank ordered according to sentence recognition accuracy scores, perceptually based speech intelligibility, and perceptually based consistency scores. Rank ordering was used to create a common metric for the scores. The speech intelligibility score was the mean perceptual rating for each participant, calculated from the responses by the 10 listeners for the Kent words across the five sessions. The speech consistency score was a coefficient of variation (CV), calculated using the mean and standard deviation ($CV = SD/M$) from the perceptual intelligibility scores across the five sessions. Pearson correlations were computed to provide information on the relationship between accuracy, intelligibility, and speech consistency. Partial correlations were calculated for recognition accuracy, intelligibility, and consistency. All three were independently used as controlling variables in calculating the partial correlations.

RESULTS

Goal #1: Recognition Accuracy

In the first session, the dysarthria group achieved mean recognition accuracy scores of 27%, compared to 39% for the control group on the Rainbow sentences. For the dysarthria group, recognition accuracy scores increased to 70% on the second session and improved to 80% by the fifth session.

Recognition accuracy scores for the control group increased to 86% on the second session and improved to 93% by the fifth session. An alpha level of .05 was used for all of the ANOVA results. An ANOVA showed that there was a significant group difference ($F[1, 10] = 10.28$, $p = .009$) in recognition accuracy scores between the dysarthria and control groups. Speakers in the control group (normal speakers) consistently outperformed the speakers in the dysarthria group (Fig. 1).

The mean recognition accuracy scores for the control group (93%) were compared to the mild (88%),

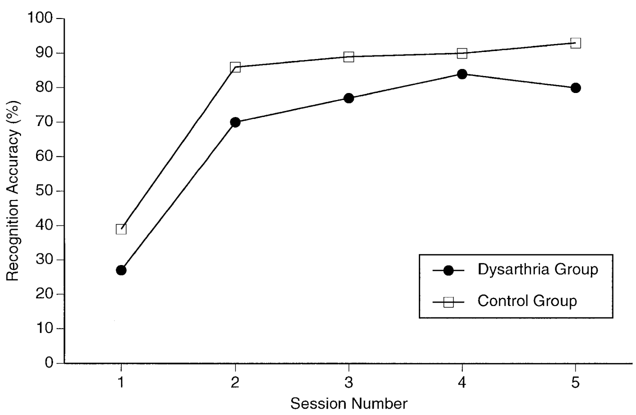

Figure 1. Recognition accuracy scores for Rainbow sentences for the dysarthria and control groups.

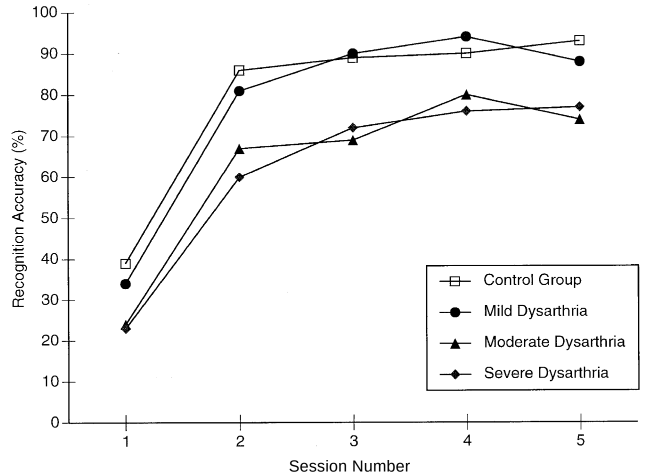

Figure 2. Recognition accuracy scores for Rainbow sentences for the subgroups of the dysarthria group and the control group.

moderate (75%), and severe (77%) subgroups within the dysarthria group from the fifth session (Fig. 2). There was no significant difference (F [1, 6] = 2.51, p = 2.51) between the control group and the mild dysarthric speakers. This may be due to the small sample size and thus a low power to detect a difference. The ANOVA revealed that the control group did score significantly higher in recognition accuracy than the moderate (F [1, 6] = 11.36, p = .015) and severe (F [1, 6] = 65.54, p = .0002) dysarthric speakers.

The acquisition curves (i.e., rate of change over sessions) for the dysarthria and control groups were compared using an ANOVA profile analysis. There was no significant difference in the acquisition curves of recognition accuracy scores between the two groups. Between the fourth and fifth sessions, the control group's recognition accuracy scores continued to improve (from 90% to 93%), while the dysarthria group's recognition accuracy scores decreased slightly (from 84% to 80%) over the same sessions. However, the difference was not significant.

Goal #2: Intelligibility/Consistency

With an alpha level of .05, the results of the Pearson correlation analysis on the Kent words revealed that there was a significant positive correlation between the 10 listeners' perceptual ratings of the dysarthria group's intelligibility and recognition accuracy scores (r = .80; p = .05). Those participants perceived as being more intelligible achieved higher scores on VoiceType. There was no significant correlation between perceptual ratings of participants' consistency of speech production and recognition accuracy scores. When a partial correlation analysis was completed controlling for consistency, the correlation between intelligibility and recognition accuracy scores was no longer significant. The magnitude of the correlation, however, remained relatively high (r = .79; p = .11). This indicates that while intelligibility is highly correlated with recognition accuracy scores, there is an interaction between intelligibility and consistency. The rankings for intelligibility, recognition accuracy, and consistency of speech production for the participants from the dysarthria group are presented in Table 2.

DISCUSSION

VoiceType's recognition accuracy was significantly greater for speakers in the control group than the speakers with moderate and severe dysarthria. There was no significant difference between the control group and speakers with mild dysarthria. The acquisition curves for the control and dysarthria groups across the first four sessions had differing elevations but were parallel and not significantly different. This result supports the findings of Ferrier et al. (1992), who concluded that the DragonDictate speech recognition system could be successfully used by speakers with mild to moderate dysarthria. Information is not yet available on whether dysarthric speakers could eventually achieve recognition accuracy levels of normal speakers given additional use of a speech recognition system, or whether acquisition curves plateau. The difference in the acquisition curves of the control and dysarthria groups between the fourth and fifth sessions suggests, however, that dysarthric speakers may plateau at lower recognition levels.

An evaluation of recognition accuracy scores for normal and dysarthric speakers producing unrelated words is presented in a paper by Doyle, Leeper, Kotler, Thomas-Stonell, O'Neill, Dylke, and Rolls (1997). These speakers were the same speakers as the participants in the current study. Comparison of results indicates that recognition accuracy scores were consistently higher for the sentences (words in context) than for the unrelated words. This likely

TABLE 2: **Participant Rankings (Dysarthria Group) for Perceptual Intelligibility, Speech Consistency, and VoiceType Scores**

Participant Code	*Severity Subgroup**	*Perceptual Intelligibility†*	*Perceptual Consistency‡*	*VoiceType Scores*	
				Words#	*Sentences§*
P1	Mild	2	1	1	1
P2	Mild	1	2	3**	3
P3	Moderate	4	6	3**	2
P4	Moderate	5	4	5	6
P5	Severe	3	5	2	4
P6	Severe	6	3	6	5

The values represent participant rankings in each category. *Severity subgroup refers to ratings of mild, moderate, or severe dysarthria of speech based on the Computerized Assessment of Intelligibility of Dysarthric Speech—word task intelligibility score. †Percent intelligibility refers to intelligibility rankings by the 10 listeners. ‡Perceptual consistency refers to speech consistency scores derived from the perceptual intelligibility scores across the five sessions. #Words refers to the single-word intelligibility test proposed by Kent et al. (1989). §Sentences refers to Rainbow sentences. **Denotes equal ranks.

reflects the higher number of multisyllabic words used in the Rainbow sentences, which provides the recognition system with more information than do one syllable words (e.g., there are more similar sounding one syllable words than multisyllabic words) and the repetition of some of the words (i.e., a, the) in the sentences, which provides the system with more opportunities to adapt. The trends for both words and sentences were similar. In both studies, the control group outperformed the dysarthria group.

There was no significant correlation between the dysarthric speakers rated perceptually as having the most consistent speech productions and highest recognition accuracy scores. Clinicians need to be aware that perceptual consistency alone is not a reliable indicator of success in speech recognition with speaker-adaptable systems. Since there is no simple way to measure acoustic consistency, clinicians should use speech recognition on a trial basis to evaluate a client's potential success with a system and avoid *a priori* assumptions based on perceptions of speech consistency. In contrast, better speech intelligibility did correlate significantly with high recognition accuracy scores. VoiceType has greater success in recognizing speech when the speaker is more intelligible. The similarities between the current study using VoiceType and the Ferrier et al. (1992) study using DragonDictate suggest that the results of this study are generalizable.

Further research is needed to determine how many sessions are required to achieve maximal adaptation of a speech recognition system for dysarthric speakers and whether dysarthric speakers can eventually achieve the same accuracy levels as normal speakers. A single-subject study with a mild dysarthric speaker has been completed to begin to address these questions (Kotler & Thomas-Stonell, 1997). It is crucial to learn more regarding which characteristics of dysarthric speech (i.e., durational cues, amplitude, or frequency features) affect recognition accuracy and whether speech training can improve recognition accuracy scores. This study takes an important first step toward determining whether perceptual measures of speech intelligibility and consistency affect recognition accuracy for dysarthric speakers. Furthermore, extensive research is needed in this area so that clinicians can reliably predict a dysarthric speaker's potential for effective use of a speech recognition system.

ACKNOWLEDGMENT

This research was supported by a grant from the Ontario Ministry of Colleges and Universities (University Research Incentive Fund) and IBM Canada Ltd. The authors wish to thank the participants and the listeners for their time and effort. The authors also wish to thank Mr. Reinhard Schuller for his assistance with the statistical analysis and Ms. R. Gannon, Speech-Language Pathologist, and Dr. M. Milner, Vice-President Research and Development of Bloorview MacMillan Centre, for their guidance and support.

Address reprint requests to: Nancy Thomas-Stonell, Thomas-Stonell Consulting, 61 Tudor Glen Cres., West Hill, ON, Canada M1C 2S3.

REFERENCES

Carlson, G. S., & Bernstein, J. (1987). Speech recognition of impaired speech. *RESNA '87*, *Proceedings of the 10th Annual Conference on Rehabilitation Technology: Meeting the Challenge, 7*, 165–167.

Coleman, C. L., & Meyers, L. S. (1991). Computer recognition of the speech of adults with cerebral palsy and dysarthria. *Augmentative and Alternative Communication, 7*, 34–42.

Doyle, P. C., Leeper, H. A., Kotler, A., Thomas-Stonell, N., O'Neill, C., Dylke, M., & Rolls, K. (1997). Comparative perceptual and computerized speech recognition functions for dysarthric speakers. *Journal of Rehabilitation Research and Development, 34,* 309–316.

Fairbanks, G. (1960). *Speech and articulation drillbook* (2nd ed.). New York: Harper and Brothers.

Ferrier, L. J., Jarrell, N., Carpenter, T., & Shane, C. (1992). A case study of a dysarthric speaker using the DragonDictate speech recognition system. *Journal for Computer Users in Speech and Hearing, 8*(1–2), 33–52.

Ferrier, L. J., Shane, H. C., Ballard, H. F., Carpenter, T., & Benoit, A. (1995). Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. *Augmentative and Alternative Communication, 11*, 165–174.

Goodenough-Trepagnier, C., & Rosen, M. J. (1991). Towards a method for computer interface design using speech recognition. *Proceedings of the 14th Annual Conference on Rehabilitation Technology, RESNA: 1991: Technology for the Nineties, 11*, 328–329.

International Business Machines Corporation. (1991). *IBM VoiceType: User's guide.* Boca Raton, FL: Author.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech Pathology, 5*(3), 7–23.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*, 482–499.

Kotler, A., & Thomas-Stonell, N. (1997). Evaluation of speech training on the accuracy of speech recognition for an individual with a speech impairment. *Augmentative and Alternative Communication, 13*, 1–10.

Schmitt, D. G., & Tobias, J. (1986). Enhanced communication for a severely disabled dysarthric individual using speech recognition and speech synthesis. *Proceedings of the 9th Annual Conference on Rehabilitation Technology, RESNA 1986: Employing Technology, 6*, 304–306.

Stern, R. M., & Lasry, M. J. (1987). Dynamic speaker adaptation for feature-based isolated word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing, 35*(6), 751–763.

Sy, B. K., & Deller, J. R. Jr. (1989). An A-I-based communication system for motor and speech disabled persons: Design methodology and prototype testing. *IEEE Transactions on Biomedical Engineering, 36,* 565–571.

Yorkston, K., Beukelman, D., & Traynor, C. (1984). *Computerized assessment of intelligibility of dysarthric speech*. Tigard, OR: C. C. Publications.

APPENDIX

Sample of Kent Word List and Foils

Kent Words		*Foils*	
bad	bed	bat	pad
sip	ship	tip	zip
spit	pit	sit	it
knot	dot	nod	nut
sigh	shy	tie	thigh
sheet	seat	feet	eat
sticks	six	ticks	stick
knew	know	knee	gnaw
leak	lick	league	reek
chair	share	tear	air

From Kent et al. (1989).