CONAN: A System for Analysis of Temporal Factors in Conversation

Norman P. Erber and John Yelland
School of Human Communication Sciences
La Trobe University

A multi-purpose laboratory apparatus (CONAN = Conversation Analyzer) has been created that collects and stores data on conversational performance. For either live or videotaped conversation, the system can automatically accumulate and display: number of speaking turns; amount of time during which each participant talked, neither talked, or both talked; and amount of time consumed by breakdown and repair. CONAN has been used in teaching and research to provide measures of normal/disordered conversational processes. Pairs of student clinicians can converse while one simulates hearing and vision impairment. A relation has been found between degree of simulated sensory impairment and amount of time required for clarification.

Several years ago, a conversational fluency rating scale was developed as a simple and efficient method for evaluating the normalcy of conversation (Erber, 1996). In practice, one observes a conversation between two people (e.g., a client and a clinician), and then assigns a number from 1 *(poor)* to 4 *(good)* or from 1 *(poor)* to 5 *(good)* to specify the degree to which the conversation appears to flow smoothly without need for clarification. When applied by experienced clinicians, such ratings are reliable both within and between raters. This rating method has been used to quickly assess the effects of various situational factors on conversational performance: simulated sensory loss (Erber, 1995), use of an amplification system (Erber & Osborn, 1994), the surrounding environment (Erber, Lamb, & Lind, 1996), and the distance between clinician and client (Erber, Holland, & Osborn, 1998).

There has been a need to extend these subjective measures of conversational performance (Duncan & Fiske, 1985). It has been hypothesized that observers

Correspondence concerning this article should be addressed to Norman P. Erber, PhD, School of Human Communication Sciences, La Trobe University, Bundoora, Victoria, Australia 3083.

typically attend to a small set of events as they assess a conversation, and consequently that conversational fluency ratings probably are related to a relatively small set of physical factors (Erber, 1996; Miller, 1956). Some likely factors are:

1. the proportion of conversation time occupied by breakdown and repair/clarification (a conversation containing considerable repair/clarification is not a fluent one),
2. the number of breakdowns (a conversation containing many breakdowns is not a fluent one),
3. the average duration of a breakdown (a conversation containing breakdowns of long duration is not a fluent one),
4. the proportion of conversation time devoted to silence (a conversation containing considerable silence is not a fluent one),
5. the proportion of conversation time that each person talks (a conversation dominated by one participant is not a fluent one), and
6. the rate of turn-taking in the conversation (a conversation that consists of an exchange of long monologues is not a fluent one).

To extend the clinician's ability to assess conversational performance, it is necessary to measure a variety of factors including those mentioned above. One measurement strategy has been to produce video tapes of conversation, carefully examine them as they are played back repeatedly, and (a) count events or (b) time the duration of events. This is a tedious method, but one that has been used previously in many clinical studies of normal and disordered language behavior (e.g., Caissie & Rockwell, 1993; Damico, 1991; Prutting & Kirchner, 1987).

More than 25 years ago, Jaffe and Feldstein (1970) developed an instrumental method for real-time measurement of conversational performance, and published a detailed study of temporal patterns in normal human interaction. Recently, we have designed and assembled a computer-based system (CONAN = CONversation ANalyzer) that also can automatically assess components of conversation as the conversation is in progress. The purpose of this study was to evaluate the usefulness of the CONAN system by simulating difficult auditory and visual perceptual conditions in people with normal hearing and vision and examining the effects on their conversational performance.

METHODS

The CONAN system contains audio and video hardware, a simple analog-to-digital converter, a computer, and specially written software. Two participants in different rooms converse by listening through headphones and watching each other on television monitors. Observers in a third room watch a split-screen video display, and analyze the interactions of the two participants (see Figure 1).

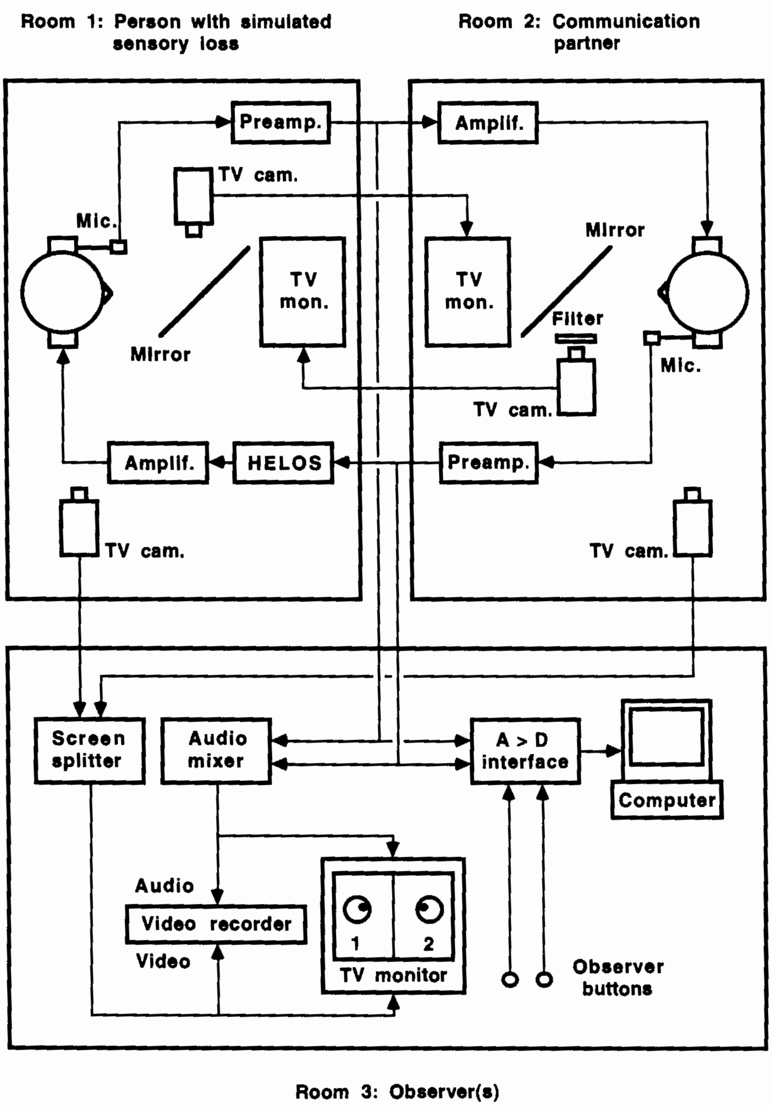

Figure 1. Block diagram of the CONAN conversation-analysis system. (Preamp = preamplifier, cam = camera, mon = monitor, mic = microphone, amplif = amplifier, A>D = analog to digital.

Hardware

CONAN is designed to monitor the vocal output of two participants. Overall, the system consists of four main parts:

(1) Audio Intercom

This part of the system allows the two participants to talk to each other. Close fitting headphones are used, with a microphone on a flexible boom that is placed beside the mouth. An independent audio path extends from each person's microphone to the other person's headphones. A HELOS hearing-loss simulator (Erber, 1995, 1996) is placed in one audio circuit so that one participant can experience simulated hearing impairment.

A suitable amount of "sidetone" is provided to the participants, as in a telephone, so that they can hear themselves naturally while speaking and thus maintain appropriate voice level. The sound level in each headphone is pre-set to about 70 dB SPL. The microphone output signals can be recorded onto a stereo audio or stereo video cassette for later analysis by CONAN or by live observers.

(2) Two-Way Video Link

This part of the system allows the two participants to see each other in a natural manner. Because the participants are in separate rooms, a visual link is provided via video cameras and monitors. A large light-proof box encloses each monitor, with a large sheet of semi-reflective glass mounted in front at a 45° angle, so that the side-mounted camera can obtain a frontal image of the participant but not impede that participant's view of the monitor (see Figure 1). This arrangement also permits correct eye-gaze direction ("eye-contact") for each participant (Acker & Levitt, 1987; Argyle & Cook, 1976).

Each video monitor is about 1.4 m from each participant. Each camera lens is adjusted so that the image of each participant's head appears life-size on the partner's monitor screen. Various filters (light-diffusing plastic disks) can be placed in front of the video camera lens to simulate vision impairment.

(3) Split-Screen Video Display

This part of the system: (a) allows external observers to watch a conversation in progress in order to rate fluency and/or measure occurrences of breakdown and repair, and also (b) permits one to make realistic video recordings of the conversation.

An additional video camera is mounted in each room – to the right of one participant and to the left of the other. These two cameras are fitted with identical lenses and fixed at the same distance and angle from each person. The furniture and background in each room are the same. Custom-built apparatus combines the two video signals, producing a split-image picture of the two participants and giving the illusion that the two people are talking to each other face-to-face without the presence of intervening equipment.

(4) Measuring Apparatus

This part of the system responds to various events during a conversation. Each person's speech signal is delivered via a preamplifier to an electronic circuit whose output closely follows its amplitude (waveform envelope). Whenever the amplitude exceeds the pre-set threshold of a comparator, an electronic switch closes to indicate that the person is speaking. The circuit's amplitude response to the beginning of any utterance is very fast (about 30 ms), but the circuit contains an adjustable time constant which slows decay – to insure that small gaps between syllables or words are ignored. Large gaps (greater than about 1.3 s) cause the switch to open (Jaffe & Feldstein, 1970). A meter indicates signal level, and an output light shows the state of the switch. A computer (AMIGA 1200) detects the state of the switch via its parallel port.

Up to six observers can operate buttons independently to log the cumulative amount of time (in seconds) consumed by repair of conversational breakdown. Whenever an observer depresses his/her button, a switch closes. The computer detects this switch closure via its parallel port.

Software

This part of the system logs the time occupied by various events during a conversation. A program was written in BASIC to interpret (a) the pattern of switch closures produced by the vocal output of the two participants during a brief conversation (e.g., 2 to 5 min) and also (b) the pattern of switch closures produced by the observers of the conversation. Switch positions are sampled 100 times per second.

(1) Talk Time

As a conversation is in progress, the computer records (in seconds): (a) the cumulative duration of participant A's vocal output, (b) the cumulative duration of participant B's vocal output, (c) the cumulative amount of time during which both participants spoke simultaneously, and (d) the cumulative amount of silence that occurred during the conversation (neither participant speaking).

The signal from each person's microphone is assumed to be speech, although background noises and human non-speech noises (e.g., a cough) also may be detected, decreasing accuracy of the measurement of talk time. A quiet background is preferred.

(2) Clarification Time

Each observer is instructed to push the button and to keep it depressed whenever conversation breaks down and the participants are engaged in conversational repair (i.e., "clarification time"). The software records the cumulative amount of conversation time (in seconds) consumed by breakdown and repair as judged by each observer *either* independently *or* in combination with other observers.

(3) Speaking Turns

A speaking turn is counted when one person fully stops speaking and the other starts to speak. If one person interrupts the other, this overlap of vocal output is not recorded as a *change* of speaking turn if the person who was interrupted continues talking.

Automatic counting of speaking turns relies on detecting the pauses between successive utterances and on a simple rule. If Person B begins to talk after Person A has stopped talking, a turn is counted. But if Person B begins to talk without waiting until Person A has stopped talking, the duration of overlap is recorded as "Both talking" and no turn change is counted. How does the computer know that Person A has stopped talking? If Person A pauses for more than 1.3 s, that person is considered to have stopped talking. This time was chosen to be greater than the typical gap between words in running speech (Jaffe & Feldstein, 1970).

Sample output data are shown in Table 1.

Table 1

Example of Data Printed Out by CONAN After a Conversation
(sens imp = sensory impairment)

Date: 10/1/95	
Person 1 (sens imp): GA	Person 2 (partner): RS
Conversation number: 14	Topic: shoes
HELOS conditions: Threshold 5; Distortion 10	
LOVIS conditions: Optical diffuser 2	
Conversation time: 2 minutes	
SENS IMP	44.5 seconds
PARTNER	40.4 seconds
BOTH:	23.3 seconds
NONE	11.8 seconds
TOTAL	120.0 seconds
CLARIF	58.1 seconds
NUMBER TURNS SENS IMP	23
NUMBER TURNS PARTNER	23

A TEACHING APPLICATION

We have begun to evaluate the potential of CONAN in teaching and research. In one brief study, four pairs of speech-language pathology students (female, age range 20-24 years) sat in separate rooms and conducted three 2-min conversations under each of 20 different combinations of simulated hearing loss and simulated vision loss (i.e., each pair conducted 60 different conversations). Simulated hearing loss was produced by a HELOS electronic hearing loss simulator (Erber, 1995, 1996) with adjustable threshold, slope, and distortion. Settings were selected that produced word-identification scores ranging from 13 to 77%. Simulated vision loss was produced by placing translucent plastic filters in front of the video camera lens; these diffused light and produced blurred images. The resulting visual conditions ranged from 6/12 (20/40) to 6/240 (20/800) visual acuity obtained with a Snellen letter chart.

One member of each pair played the role of a client with impaired hearing and vision, while the other person played the role of a partner with normal hearing and vision. For each conversation, the conditions of sensory loss (e.g., 6/60 visual acuity; threshold = setting 4, distortion = setting 8, sloping audiogram) and the topic to be discussed (e.g., familiar subjects such as cats, garden, football) were selected randomly from prepared decks of cards. Two student clinicians (female, ages 24 and 27 years), experienced in speech/language assessment, played the role of observers. They watched the conversations on a television monitor in a separate room, and each pressed a response button during periods of communication breakdown and repair. The two buttons were in series, and so *agreement* between the two observers was measured. As each conversation was in progress, CONAN measured: (a) temporal speech output data and (b) time of observed breakdown and repair (clarification).

RESULTS/DISCUSSION

Figure 2 shows the amount of conversation time that was occupied by clarification under each of the 20 combinations of simulated hearing loss and vision loss (averaged over 4 pairs of communicators and also three 2-min conversations). When vision is virtually normal (6/12 acuity), the participants can maintain nearly normal face-to-face conversation regardless of degree of hearing impairment. Under these conditions, the clients can rely on lipreading and other visible facial cues to compensate for their loss of auditory detail. When their vision is degraded (from 6/60 to 6/240 acuity), however, more active repair is required to maintain conversation, especially when they experience greater hearing impairments.

These results are reflected in clinical experience. Often an older person (and the person's family) will not notice a gradually progressive loss of hearing, because the person is able to unconsciously compensate through lipreading, and/or the family may unconsciously compensate with increased clarity of speech. When the person's vision also begins to diminish, however, these simple compensatory strategies may no longer be sufficient, conversational performance deteriorates, and the hearing loss becomes apparent.

Figure 3 details the results of the fourth pair of student communicators described above ("G" and "H"), selected to show how strongly conversational behavior might be influenced by sensory loss. The number of seconds occupied by clarification (out of 360 total seconds of conversation) is shown as a function of the proportion of time Communicator G (the client) talked during the conversa-

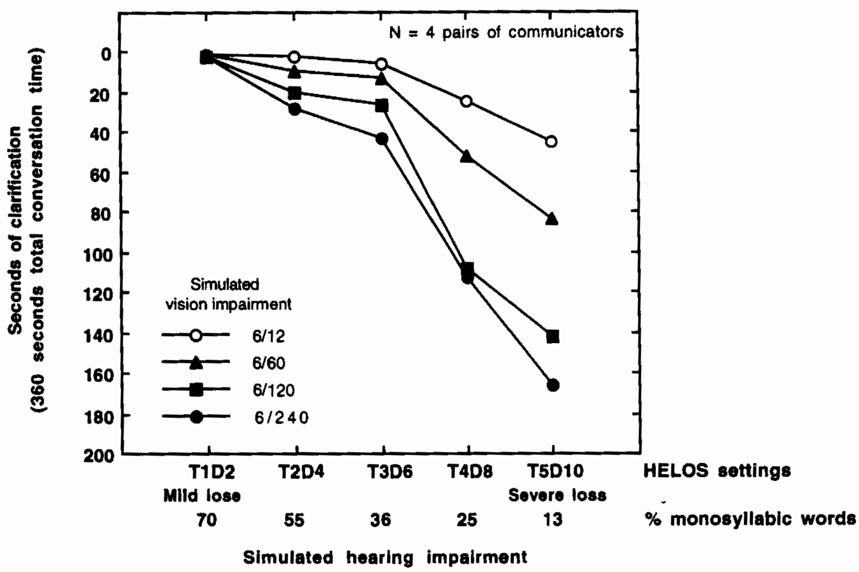

Figure 2. Effects of simulated hearing and vision impairment on amount of clarification in conversation. Data are averaged over 4 pairs of communicators and three 2-min conversations.

tion (see Erber, 1996, pp. 204-206). The four sets of symbols that are shown represent data accumulated under each of the four different conditions of simulated vision loss. As in Figure 2, the lines that connect the points trace a progression through five different conditions of simulated hearing loss. Thus four different possible aging scenarios are shown for Communicator G: four different amounts of acquired vision loss, each combined with a gradual change in acquired hearing loss from mild to severe.

When Communicator G has normal vision (6/12 acuity; open circles), relatively little clarification is needed in conversation, and moreover, G speaks only about 25-45% of the time – regardless of the amount of her hearing loss. This proportion of talk time is typical of her normal everyday interaction with H. When G experiences reduced vision as the result of the other three simulated conditions, more conversational clarification is required, and G's proportion of talk time increases. Under the most difficult conditions of combined hearing and vision loss, G even begins to *dominate* conversational talk time (>50%), which is not typical of her normal daily interaction with H.

In our clinical experience, some older people demonstrate what appears to be a "personality shift" as they age. They seem to become preoccupied with themselves, and seem to talk excessively during conversation. They may simply be dominating conversation to avoid misperception of the speech of others. This is

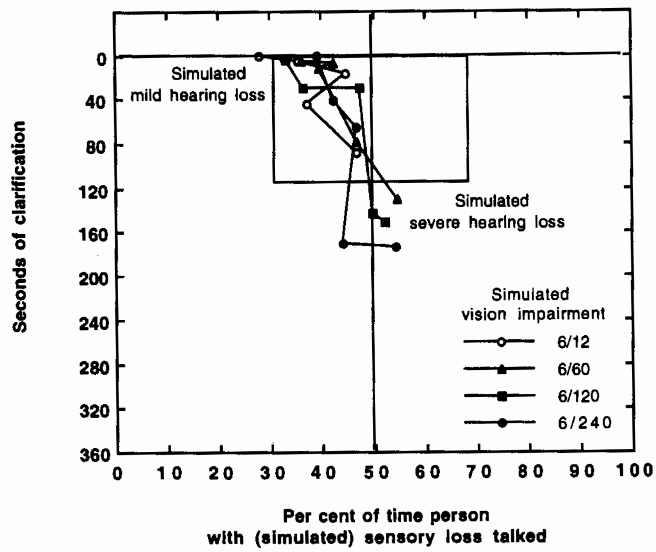

Figure 3. Number of seconds of clarification as a function of the per cent of time the person with (simulated) sensory loss talked. Data are presented for one pair of communicators.

one of several (mal)adaptive approaches to sensory loss. Such people may be unaware of more effective strategies that they can use (Erber, 1993).

SUMMARY

The CONAN system appears to be a reliable and efficient method for data acquisition. A variety of descriptive information is available immediately after each conversation. No repetitive videotape analysis is required, although a more detailed approach may still be necessary for an analytical study of conversational performance. Video recordings of conversation may be replayed repeatedly to examine the reliability of equipment or observer judgements/ratings. Video recordings also may be replayed to study communicator behaviors other than overt instances of breakdown and repair. Different instructions may be given to observers to modify their attention and response criteria (e.g., "Press the button whenever the participant on the left leans forward – to hear or see better.").

There has long been a need to examine real conversations by real clients in real conversational settings. We believe that video recordings obtained in locations external to the laboratory also may be analyzed by CONAN, but only under the following special conditions: (a) a video recorder with *stereo audio* inputs is used to record the conversation, (b) background noise is minimized, and (c) sufficient acoustic channel separation between the participants is maintained. Good lighting and lapel microphones may be necessary.

The CONAN apparatus has the potential to become a powerful tool for research – to study effects on conversation of various factors, for example, a client's degree of sensory loss, poor communication environments, topic familiarity, or partner awareness and cooperation.

In conjunction with simulation of hearing and vision impairment, the apparatus also has the potential to become part of a comprehensive learning laboratory for university students. Students can recreate the sensory aspects of aging (through a progression of increasingly difficult simulations) and experience the combined effects on communication. Results that have been obtained from students simulating sensory loss in this way are consistent with the reports of real clients.

REFERENCES

Acker, S.R., & Levitt, S.R. (1987). Designing video conferencing facilities for improved eye contact. *Journal of Broadcasting and Electronic Media, 31,* 181-191.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze.* Cambridge, England: Cambridge University Press.

Caissie, R., & Rockwell, E. (1993). A videotape analysis procedure for assessing conversational fluency in hearing-impaired adults. *Ear and Hearing, 14,* 202-209.

Damico, J.S. (1991). Clinical discourse analysis: A functional approach to language assessment. In C.S. Simon (Ed.), *Communication skills and classroom success* (pp. 125-148). Eau Claire, WI: Thinking Publications.

Duncan, S., & Fiske, D.W. (1985). *Interaction structure and strategy.* Cambridge, England: Cambridge University Press.

Erber, N.P. (1993). *Communication and adult hearing loss.* Clifton Hill, Australia: Clavis Publishing.

Erber, N.P (1995). Applications of hearing loss simulation in the education of student clinicians. *Journal of the Academy of Rehabilitative Audiology, 28,* 37-50.

Erber, N.P. (1996). *Communication therapy for adults with sensory loss* (2nd ed.). Clifton Hill, Australia: Clavis Publishing.

Erber, N.P., Holland, J., & Osborn, R.R. (1998). Communicating with elders: Effects of distance. *British Journal of Audiology, 32,* 135-138.

Erber, N.P., Lamb, N.L., & Lind, C. (1996). Factors that affect the use of hearing aids by older people: A new perspective. *American Journal of Audiology, 5,* 11-18.

Erber, N.P., & Osborn, R.R. (1994). An alternative use for BTE FM listening systems. *Hearing Instruments, 45,* 19-21.

Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialog.* New York: Academic Press.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63,* 81-97.

Prutting, C., & Kirchner, D. (1987). A clinical appraisal of the pragmatic aspects of language. *Journal of Speech and Hearing Disorders, 52,* 105-119.