

MECHANISMS OF HEARING

Edited by

William R. Webster

Department of Psychology

Monash University

and

Lindsay M. Aitkin

Department of Physiology

Monash University

MONASH UNIVERSITY PRESS

Clayton, Victoria, Australia 3168

1983

COMPRESSION SYSTEMS FOR COCHLEAR PROSTHESES

M.W. White, Department of Otolaryngology, University of California, San Francisco, CA 94143, U.S.A.

In any practical cochlear prosthesis, the wide dynamic range of our auditory environment must be "compressed" or "mapped" into the very narrow operating range of the implant recipient. To design effective compression systems, the strong interrelationships between the temporal and spectral characteristics of auditory signals and compression systems must be fully appreciated. There are three fundamental factors in the design of compression systems for speech: (1) the statistical characteristics of the speech signal at the compressor's input; (2) the characteristics of the "transmission channel" at the compressor's output, such as the channel's operating range and the distribution of intensity differences limens across this dynamic range; and (3) the spectral-temporal fidelity necessary for the successful transmission of perceptually significant information. These factors are highly interdependent. Most difficulties in the design of compressors arise because the input signal has a relatively wide-band spectrum in which the signal level varies rapidly in time. The design problem is further aggravated if the input signal's operating range is very much larger than the output channel's dynamic range. Figure 1 is a block diagram of a typical feed-forward compressor. The designer might design a compressor in which the envelope detector's integration window is relatively long in duration (i.e., at least as long as one cycle of the lowest expected frequency component). In so doing, the designer hopes to reduce any spectral

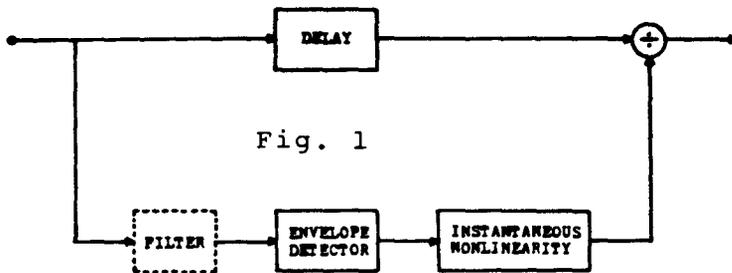


Fig. 1

distortion that might be generated by the compressor over the entire spectral range. However, if the input signal is relatively broad-band, relatively rapid changes in signal level can occur. Because the integration window is long in duration, the compressor's gain will only very sluggishly respond to these relatively fast level changes. As a

consequence, the compressor's output level will also rapidly change by approximately the same ratio as the ratio change at the input. In other words, very little compression will occur during such brief intervals. The designer has sacrificed "temporal fidelity" for spectral fidelity. At the other extreme, the designer could choose to use a very short duration integration window to improve the rate at which the compressor's gain can change, and thereby allow the compressor to appropriately compensate for rapid changes at the input. However, this approach can generate a large amount of spectral distortion particularly when relatively low frequency spectral components are present. Both harmonic distortion and intermodulation distortion products will be generated with such short duration envelope detector windows.

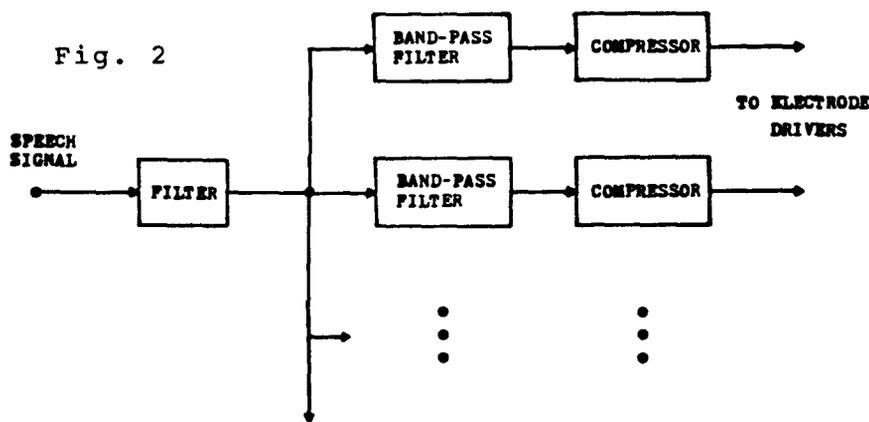
Up to now, we have considered only broad-band input signals. If a wide-band signal (e.g., speech) is filtered prior to compression by a relatively narrow-band filter, the filter will limit the risetime of the envelope of the signal. For example, if the envelope of the input signal is a step, the envelope of the filter's output will be similar in shape to the integral in time of the envelope of the filter's impulse response. Under these narrow-band signal conditions, a compressor can be designed with very good temporal and spectral characteristics, where temporal and spectral distortions are kept to a minimum. Under narrow-band conditions, the signal's level will change at a relatively slow rate compared to the duration of a cycle of any spectral component within this narrow band of frequencies. As a consequence, the compressor's envelope detector can utilize integrating windows of sufficient duration so as not to significantly distort the spectrum of these narrow plane signals. Such relatively long integrating windows will still enable the compressor to adjust its gain quickly enough to "compress" the relatively slow changing signal levels of the narrow-band signal. In designing low distortion compressors for wide-band signals such as speech, one might use a bank of filters and compressors, each filter-compressor pair individually matched in transient performance.

Spectral information above 250-1000 Hz may not be accessible to the CNS with single channel electrical stimulation although these higher frequencies are audible. Although auditory nerves phase-lock to electrical stimuli up to at least 3.2 kHz (Loeb, et al. 1983), psychophysical measures (Merzenich, 1973; Simmons, 1981; Tong & Clark, 1983) indicate that: a) the pitch percept does not

change significantly for stimulus frequencies above 250-1000 Hz; and b) frequency discrimination is nearly nonexistent above 250-1000 Hz. As a consequence, multichannel processors have been designed to transmit this higher frequency spectral information by differentially exciting different sectors of the auditory nerve. Encouragingly, there is evidence that pitch is at least a weak function of the stimulating electrode's location and that useful discrimination between stimulated electrodes is possible (Tong & Clark, 1983; Eddington, 1978). In the following discussion, spectral information that is represented by differences among the channel output levels will be referred to as "across-channel spectral information". Spectral information that is represented in the fine-grain temporal patterns within given processor channels will be defined as "within-channel spectral information".

One proposed multi-channel speech processor incorporates a channel vocoder-like speech processing system in which the output of each band-pass filter is used to excite a different sector of the nerve. How should the output of each filter drive a restricted sector of the auditory nerve? We could model each sector of nerve as a noisy transmission channel which has a very restricted operating range. In one approach, the highest signal level at a band-pass filter's output would be mapped to the top of the transmission channel's dynamic range and the lowest signal level that is to be audible would be mapped to the bottom of the channel's dynamic range. Intermediate signal levels would be mapped according to the perceptual significance of the intermediate levels and transmission characteristics of the channel. For example, intermediate levels could be mapped in accordance with the distribution of intensity difference limens across the dynamic range.

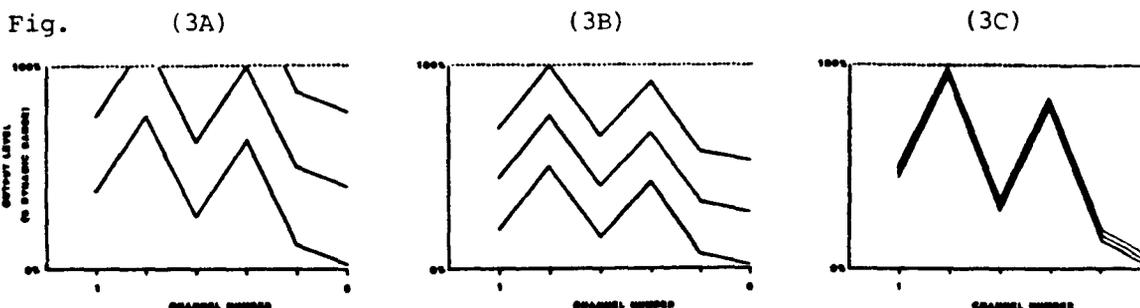
Compressors are commonly used to perform this type of intensity mapping function. One processing method is illustrated in figure 2. The initial filter



block represents the spectral pre-emphasis function in which the higher frequencies of speech are boosted to approximately the same amplitude, on the average, as the lower frequency components. Because no nonlinear function is introduced

before the filter bank, across-channel spectral distortion is not generated before the filter-bank. Because each channel is assigned its own compressor, each channel's compression function can be individually contoured to the input and output characteristics of each individual channel. As previously suggested, each compressor's transient behavior can be individually "matched" with the transient response of that channel's bandpass filter. In this manner, within-channel spectral and temporal distortion can be minimized.

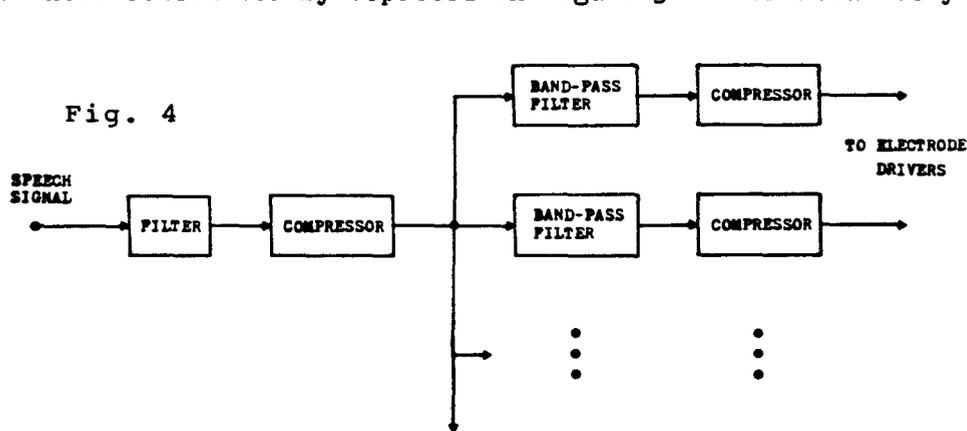
The graphs in figure 3 illustrate channel output levels of hypothetical six channel processors responding to a "steady-state vowel" stimulus delivered at three stimulus levels. Each of the three curves in each graph represents the output levels to a different input signal level.



The three representative input levels illustrate the range of absolute level

variations due to differences among speakers, differences in the speaker's distance from the microphone, differences in "loud" speech versus whispered speech, and differences in the levels within a given speech utterance (e.g., vowels are generally higher in level than are consonants). Each channel's output level is plotted as a percentage of dynamic range. A disadvantage of the processor of figure 2 is illustrated in figures 3A and 3B. For two of the three stimulus levels in figure 3A, some of the channels would be perceived as uncomfortably loud. By compressing the channel signals more than in example 3A, none of the "hypothetical" nerve sectors or "channels" are stimulated above their maximum comfortable loudness level (MCL) nor below their thresholds (see figure 3B). The large amount of compression necessary in 3B can significantly reduce the amount of across-channel spectral information available to the auditory system, particularly because only a relatively small number of intensity difference limens are available across the dynamic range of most implant recipients. This reduction in the amount of spectral information available is represented in figure 3B by the reduced differences in the channel output levels compared to those in figures 3A and 3C. Because the channel output levels form a pattern that represents the signal's spectrum, any reduction in the differences between channel output levels can reduce the transmission of spectral information to the CNS. An improvement in the signal-to-noise ratio of the across-channel spectral representation is likely to be particularly important during the brief, transitional segments of speech.

How can we improve the transmission of across-channel spectral information and also insure that the channel output levels will not go above the upper limit or below the lower limit of each channel's dynamic range. We could "decompress" the across-channel spectral representation of 3B and utilize nearly the full dynamic range of the transmission channels for the representation of spectral information (eg, see figure 3C). To a first approximation, the processor in figure 4 is capable of generating spectral representations similar to those schematically depicted in figure 3C. For relatively large changes in



the input level, the initial compressor's output level will change relatively little. The degree of compression controls exactly how much the overall output level changes for a given change in the input level.

This initial compressor reduces the proportion of transmission channel resources (ie, the auditory nerve "channel" resources) that are devoted to absolute intensity information and reassigns those resources to the transmission of spectral information which is represented by the differences in the outputs of the processor channels (e.g., figure 3C).

The amount of compression generated before the filter bank, controls the trade-off between the representation of across-channel channel spectral information and absolute intensity information. The filtering stage, placed just ahead of the first compressor, controls the relative suppressive power of the individual spectral components of the input signal. Those spectral components that are highest in level at the output of the initial filtering stage will have the most effect in "attenuating" or "suppressing" the other spectral components because these higher-level components will be dominant in determining the gain of the initial compressor. The compressors located "after" the filters simply "map" the range of levels at the individual filter outputs into the acceptable range of levels at the individual electrodes.

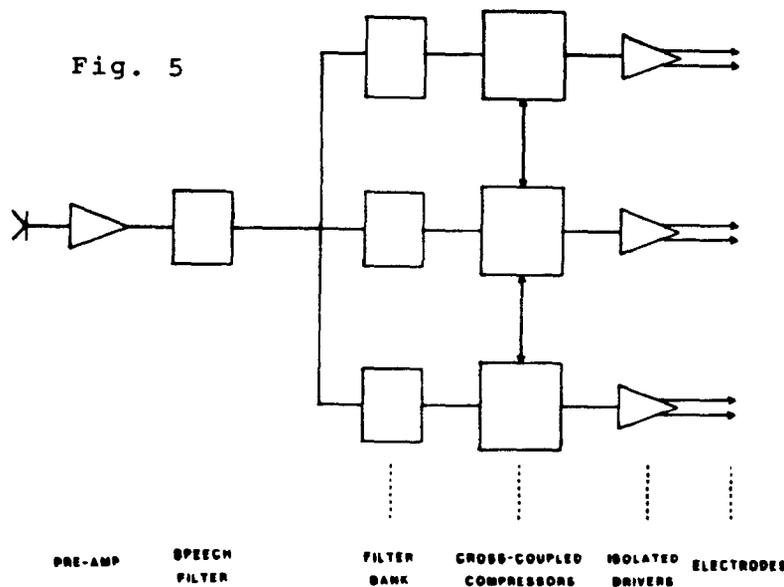
Psychophysical measures in normal hearing and in implanted subjects of "intensity discrimination vs stimulus intensity" could be used in specifying midrange mapping functions. If the implanted subject is not sensitive to within-channel spectral distortion, these post-filter compressors could be simply instantaneous nonlinearities. On the other hand, non-instantaneous post-filter compressors might be useful in performing certain temporal transforms that could improve speech reception. For example, they might be used to mimic certain perceptually relevant features of adaptation (Smith, 1983).

The formant frequencies, the shape and width of the spectral peaks, and the diffuse-compact spectral cues are critical in the perception of speech by normal hearing individuals (Stevens, 1983). However, it's possible that the sacrifice of some absolute intensity information could cause a decrease in the subject's speech reception ability over and above that gain realized from improving the spectral representation. However, data from two separate studies on compression and its usefulness in aiding the hearing impaired (Lippmann et al, 1981; Henrickson, 1982) support the relative importance of spectral information. In both studies, compressors which emphasized the signal's spectral content as apposed to the signal's absolute intensity (i.e., compressors analogous to the one depicted in figure 4) performed better in speech tests than did processors similar to the one illustrated in figure 2. In our work with a subject implanted with a multi-channel scala tympani electrode array, we have obtained evidence that also supports the value of emphasizing across-channel spectral information (White, 1983). Only two basic processor configurations were studied. Over the relatively restricted range of parameters that were examined, a processor configuration similar to that in figure 4 performed considerably better in speech tests than did a processor with a configuration similar to that in figure 2.

A significant amount of absolute intensity information is still present in the spectral representation of figure 3C, although not nearly as much information as in other models (e.g., see figures 3A or 3B). Although each channel contains relatively little absolute intensity information, the "sum" or "average" of the response changes across the channels could offer a relatively low noise representation of the input signal's absolute intensity, particularly if a large number of channels are used in obtaining the estimate.

Up to now, we have considered only an idealization of the performance of the initial compressor in figure 4. We temporarily ignored any spectral-temporal distortion that this compressor is capable of generating. Unfortunately, the initial compressor in figure 4 has the same difficulties as other single channel compressors which receive wide-band, rapidly changing input signals. For example, if the processor in figure 4 requires relatively fast compression (e.g., in order to quickly compress rapid and large amplitude excursions) the initial compressor in figure 4 can significantly distort the signal's spectrum. The signal's spectrum could be distorted to such an extent that the benefits gained by reallocating the channels' resources may be significantly diminished. If relatively slow compression is used spectral distortion can be reduced, but at the expense of generating abnormally large and rapid amplitude excursions at the compressor's output (i.e., "temporal distortions").

The processor illustrated in figure 2 does not generate any across-channel spectral distortion but at the same time, it does not have the potential for emphasizing the across-channel representation of the signal's spectrum. The

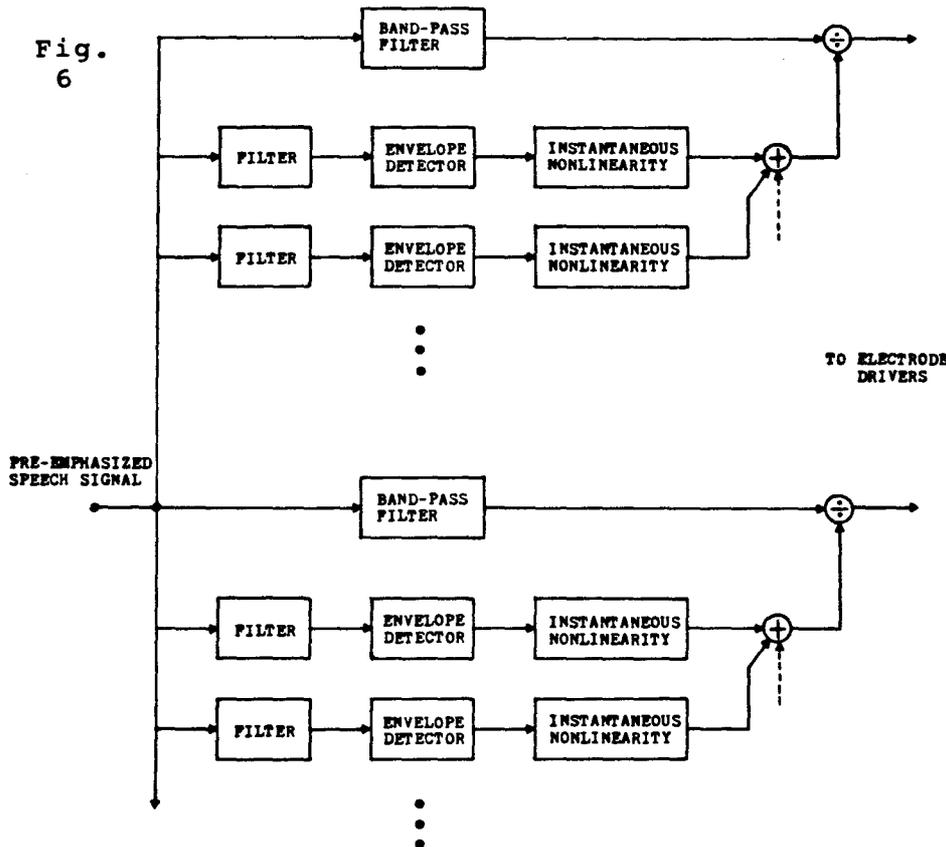


processor in figure 4 is capable of emphasizing across-channel spectral information, but the pre-filter-bank compressor can distort the signal's spectral representation. Other processors offer better temporal and spectral characteristics. Figure 5 is a diagram of the basic "cross-coupled" configuration. In this configuration, each compressor's gain is controlled by a weighted combination of the signal levels derived from some or all of the channels. The cross-coupled configuration does not generate spectral distortion at any point before the spectral analysis stage

(i.e., the filter bank) and it allows the emphasis of across-channel spectral information. Within-channel spectral and temporal distortions are minimized if

each compressor's transient response (which includes the contributions of those channels which affect the gain of that channel) is appropriately matched to the transient response of that channel's bandpass filter. Figure 6 illustrates a processor similar in concept to that illustrated in figure 5. In figure 6, each channel's gain is independently derived from a weighted sum of signal levels derived from one or more frequency ranges. For example, if a filter preceding one of the envelope detectors has a high gain for a certain band of frequencies, this band will be particularly capable of attenuating that channel's output amplitude. The temporal features of the inter-channel interactions can also be precisely controlled. This configuration is even more versatile if an instantaneous nonlinearity is placed between the "summer" and "divider" stages.

To emphasize across-channel spectral information, each channel's gain should be controlled by a signal which is an estimate of the input signal's intensity. An estimate of the input signal's intensity can be obtained from: (a) the sum of the envelope signals from all bands; (b) the sum of the envelope signals from spectral components within and moderately near the channel's passband; (c) the sum of the envelope signals from spectral components within and lower than the channel's pass-band. The peripheral auditory system appears



to use a weighted combination of "b" and "c". If spectral and temporal distortions are to be minimized, each compressor's transient response should be appropriately matched to the transient response of that channel's bandpass filter. As a consequence, the transient characteristics as well as the static characteristics of these across-channel interactions must be appropriately controlled. As with single channel compressors, the "gain-controlling"

signal cannot change rapidly during a cycle of any spectral component within that channel. At the same time, the gain-control signal must change quickly enough to compensate for level changes in these same spectral components. These conditions constrain the design of such compressors. As a specific example, a very fast, high-amplitude transient within a higher frequency channel could very quickly reduce the gain of the lower frequency channels and produce a sudden "notch" in the output waveforms of these low frequency channels. Such short duration "notches" can generate severe spectral distortions within the low frequency channels. This type of distortion can be minimized by restricting the coupling between the channels, such that the high frequency channels cannot reduce the gain of the lower frequency channels at too high a rate relative to the cycle duration of those lower frequency channels. This can be accomplished in at least two ways. The rate at which the gain control signal changes can be constrained by using a longer duration integrating interval for controlling the gain of the lower frequency channels. The integrating interval should be at least one to several periods of the frequency components within the channel whose gain is to be controlled. In other words, the rates of change of the

"gain-controlling" signals should not be fast compared to the cycle durations of the signals within the pass-band of the channel. Another method for reducing such spectral distortions would be to simply disallow relatively fast changing gain control signals from affecting the gain of those channels which contain spectral components with relatively long cycle durations. Indeed, the peripheral auditory system does not generate suppression at frequencies significantly lower than the suppressor's frequency.

What improvement in the transmission of the signal's spectrum could we expect? First, let us assume that the speech has been appropriately pre-emphasized. From each of the channel filters, we would expect nearly equal output levels if we averaged over a very large sample of speech. Also, let us assume that for most brief segments of speech (ie, segments on the order of 10-20 msec), the maximum difference between the output levels of the channel filters will be no greater than about 20-30 dB (Stevens, 1983). Furthermore, we assume that the processor should be capable of compressing at least a 60 to 90 dB absolute intensity range of speech and environmental signals. Under these assumptions, the multichannel processors illustrated in figures 5 and 6 could offer a factor of 2 to 4.5 improvement in the signal to noise ratio of the across-channel spectral representation compared to that generated by the processor illustrated in figure 2.

With the compression systems illustrated in figures 5 and 6 it is possible to "expand" or "contract" the across-channel spectral representation. In one extreme example of this expansion, only those channels "centered" at the spectral peaks (e.g., at the formant frequencies of the speech signal) would receive suprathreshold stimulation. Such a compression system could be implemented by using a large amount of cross-coupling between "nearby" channels. The cross-coupling would be very highly "inhibitory". The channel receiving the largest amplitude signal would effectively "turnoff" all of the other channels which are strongly and "negatively" coupled to this channel. This type of processing is similar to that used in peak-picking channel vocoders described by Peterson and Cooper (1957). This type of processing is also very analogous to the visual edge enhancement process first described by Mach.

- Eddington, D. (1978) Annals of Otolaryngology, Rhinology, and Laryngology (Suppl. 38) 86:1-76.
- Henrickson, L. (1982) Doctoral Dissertation, Stanford University.
- Lippmann, R., Braida, L., Durlach, N. (1981) Journal of the Acoustical Society of America, 69:524-534.
- Loeb, G., White, M., Jenkins, W. (1983) Annals of the New York Academy of Sciences, 405:79-93.
- Merzenich, M., Michelson, R., Pettit, C., Schindler, R., Reid, M. (1973) Annals of Otolaryngology, Rhinology, and Laryngology, 82:486-503.
- Peterson, E. Cooper, F. (1957) Journal of the Acoustical Society of America, 29:777.
- Simmons, F., White, R., Mathews, R., Walker, M. (1981) Annals of Otolaryngology, Rhinology, and Laryngology, 90:15-18(suppl.).
- Smith, R. (1983) Annals of the New York Academy of Sciences, 405:79-93.
- Stevens, K. (1983) Annals of the New York Academy of Sciences, 405:2-17.
- Tong, Y., Clark, G. (1983) Annals of the New York Academy of Sciences, 405:264-267.
- White, M. (1983) Annals of the New York Academy of Sciences, 405:348-359.