Veterans Administration

Journal of Rehabilitation Research and Development Vol. 24 No. 4
Pages 119–126

A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation

D. GRAUPE*, J.K. GROSSPIETSCH and S.P. BASSEAS****

Department of Electrical Engineering and Computer Science, Department of Biomedical Engineering, University of Illinois at Chicago; Intellitech, Inc., Northbrook, Illinois

Abstract—This paper discusses a single-microphone-based self-adaptive filter of environmental noise from speech. This filter, based on the work of Graupe (3) and of Graupe and Causey (4), has been incorporated in standard in-the-ear (ITE) and in behind-the-ear (BTE) hearing aids by several hearing aid manufacturers. Intelligibility tests by the authors and by independent researchers are presented in this paper to illustrate the filter's preformance. Significant monosyllabic-word-list intelligibility improvements are shown in hearing-impaired and in normal-hearing subjects for virtually any environmental noise, including white noise, babble (interfering background conversations), cafeteria noise, high-frequency noise, and low-frequency noise at signal-to-noise ratios to below −20 dB.

INTRODUCTION AND BACKGROUND

Although the problem of speech intelligibility in the presence of noise affects any person functioning in a noisy environment, the noise problem is more severe for those with impaired hearing. The noise problem is, and has repeatedly been cited as, the number one complaint by hearing aid users (12). The hearing-impaired person, who has difficulty in understanding speech in a noise-free environment, has even more difficulty when environmental noise is present. The discrimination capabilty of the hearing-impaired person is already diminished even for audible signals, thus he or she cannot discriminate between sounds of speech and noise that are similar. As a result, masking by noise affects the hearing-impaired person to a greater degree than it does the normal-hearing person. Further, hearing aids amplify noise and speech indiscriminately.

There are many approaches to speech enhancement in the presence of noise. Most of them are nonadaptive and do not discriminate between speech and noise. The most common nonadaptive approach is that of reducing the gain of the hearing aid at frequencies below 500 or 700 Hz. This assumes that many environmental noises are of low frequency, which is often the case. However, this approach equally suppresses low-frequency speech cues and it does so regardless of the presence of noise. The hearing-impaired person, who typically has the greatest hearing loss in the high frequencies, is now also deprived of low-frequency cues. Meanwhile, high-frequency noise is unaffected.

Another nonadaptive approach is that of using AGC (automatic gain control), a form of amplitude compression. This approach also does not discriminate between speech and noise. It suppresses both speech and noise to an equal degree, but only if the total level of either or both is above a threshold value.

There are also several self-adaptive methods for the filtering of noise from speech. The most successful is based on Widrow's filtering theory (15), and on the rather similar Tsypkin theory (13). The

*Dept. of Electrical Engineering and Computer Science and Dept. of Bioengineering, University of Illinois at Chicago, IL 60680; Vice President, Intellitech Inc., 900 Skokie Blvd., Northbrook, IL 60062.
**Intellitech Inc., 900 Skokie Blvd., Northbrook, IL 60062.

Widrow-Tsypkin approach, while rigorous, is not totally adaptive in that it requires prior parameter knowledge of speech alone or of noise alone. One approach to this problem is to use two microphones, one being directional to the noise source alone. This requirement limits the practical usefulness of the technique in hearing aid applications. Its performance is good down to a SNR (signal-to-noise ratio) of −12 dB, yielding suppression of noise by up to 10 dB (2). Its performance, however, worsens considerably if the noise comes from the direction of the speech, as in a lecture hall or similar situation.

Among single-microphone adaptive filtering approaches is that of Weiss and Aschkenasy (14). This technique uses a cepstrum-like transformation and signal reconstruction. This method improves the SNR in white noise, but with little improvement in intelligibility partly due to poor reconstruction of unvoiced speech (6). It involves time delays for four FFT calculations (at least 512 points each), which is a major obstacle to hearing-aid applications.

Boll (1) reports similarly poor intelligibility enhancement for suppression of noise using direct spectrum subtraction of averaged speech spectra.

The method of Sambur (10), where the input signal of speech-plus-noise is delayed through a tapped delay line, is essentially a single-microphone extension of the Widrow-Tsypkin approach. It is assumed that the speech is highly correlated in contrast to noise and, thus, the correlated part of speech plus noise is considered to be speech in order to provide the speech parameters for subsequent filtering. This philosophy is inadequate for unvoiced speech, which is stochastic in nature compared with voiced speech. It also fails when noise is correlated, as it often is. It is not surprising then that this method also does not improve intelligibility (2). An improved version of Sambur's method, in which the high frequencies are enhanced (2), is reported to improve intelligibility, though only for signal-to-noise ratios of at least +6 dB.

It is not surprising that Lim and Oppenheim conclude that " . . . However, almost all of these systems in fact reduce intelligibility . . . " (9), and later Lim states: " . . . Various speech enhancement systems discussed . . . appear to improve speech quality but not speech intelligibility. Intelligibility improvement when the degradation is due to wideband random noise or speech-like noise, . . . requires a fresh new approach . . . " (8) (speech-like noise being, as we understand, babble-type noise, i.e., interference due to several background conversations). Lim also claims: " . . . no algorithm has been shown to improve speech intelligibility when speech is degraded by wideband random noise and there is only one microphone input. In the case of interference from competing speakers, improvement has not been demonstrated in either intelligibility or quality." (7) Furthermore, he continues: "Even with multiple microphone inputs, significant intelligibility improvement has been demonstrated only in restricted environments." (7)

It is when intelligibility improvement is sought under a wide variety of realistic noise situations that the single-microphone self-adaptive filter to be described is applied.

DESCRIPTION OF THE ZETA SELF-ADAPTIVE FILTER

Principles

The Zeta noise blocker (first patented in 1976) is a single-microphone self-adaptive filter whose block diagram is given in **Figure 1**. The input to the filter is y(t) where $y(t) = s(t) + n(t)$: t = time, s(t) and n(t) denoting speech and noise. It operates in real time such that no delay exists between s(t) and the corresponding filtered speech output $\hat{s}(t)$.

The filter first identifies the parameters vector $\underline{p}(y)$ of the input y(t). Subsequently, parameter separation is performed, where the parameter vector $\underline{p}(y)$ is separated in sub-vectors of noise parameters and of speech parameters. The later parameter separation, as is justified theoretically in Section 11.2 of (3), is possible only when assuming that speech parameters have different features than noise parameters. Therefore, only signals whose parameter features are different from speech parameter features will be identified as noise and thus be filtered. This is the situation for most environmental noises, since their spectral or time-domain parameters vary slowly in time relative to speech parameters. Indeed, speech phonemes have characteristic random-like changes, at intervals of 5 to 200 milliseconds, depending on their being voiced or unvoiced. Hardly any environmental noise displays such time variations. Even speech babble does not have the same characteristic time variation because

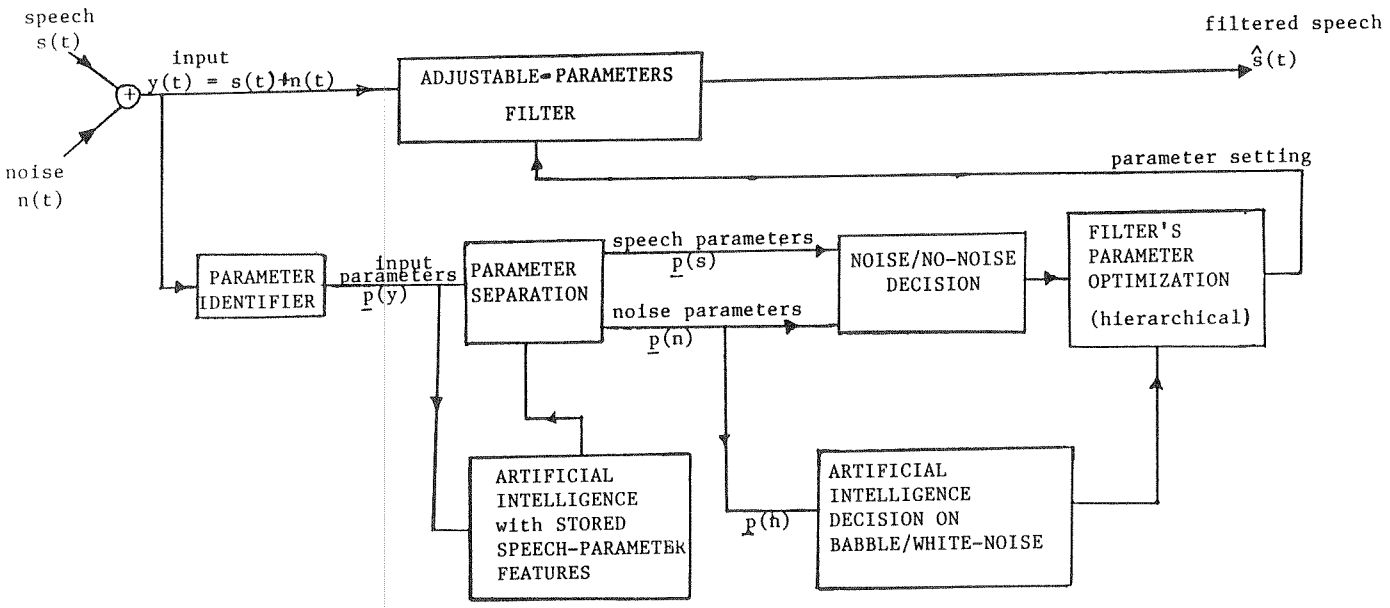

Figure 1.
Schematic diagram of Zeta self-adaptive filter.

of the acoustic averaging effect over many simultaneous conversations.

In Section 11.2 of (3) it is shown that even a vague knowledge of model order suffices for parameter discrimination. In the present realization, an artificial intelligence that considers features of speech parameters in speech recognition algorithms is employed.

Once parameter separation has been performed, a decision on the existance of noise n(t) in the input y(t) is possible. When no noise has been found to exist, the filter becomes transparent, such that the filter's output ŝ(t) is set to be equal to y(t). Otherwise, the filter (which is essentially an adjustable Wiener filter) is automatically tuned to optimally filter out the noise that has been identified. Although the filter is a linear filter, its optimization differs from that of the Wiener filter. The difference is that it employs an artificial intelligence to heuristically modify the original Wiener filter's optimization by taking into account characteristics of speech phonemes: see Chapter 10 of (3). This artificial intelligence also facilitates heuristic adjustments for babble and white noise conditions, as identified by the filter (5).

Realization

The Zeta noise blocking adaptive filter is presently realized in a 0.219 × 0.159 inch CMOS custom chip which includes an A/D (Analog to Digital Converter) on board.[1] It is small enough to fit into any conventional BTE (behind-the-ear) hearing aid and a large percentage of ITE (in-the-ear) hearing aids. It uses standard hearing aid batteries (nominal 1.4 Vdc), including Zinc-Air and rechargeable batteries, as is required for hearing aid applications. Its power consumption is approximately 1 mA; thus it adds about one third to the current drain rate of the battery.

A more powerful version of the same filter for non-hearing-aid applications also has been constructed and tested by the present authors. This is the CAF (Communication Adaptive Filter). It employs the same principles and essentially the same algorithms. However, it is not limited in its dynamic range to a 1.4 Vdc supply, nor is its size restricted to a single CMOS chip of the Zeta's present dimensions. The Zeta self-adaptive filter itself is capable of working with a 3.0 Vdc supply and will yield somewhat better performance than with a 1.1 Vdc to 1.3 Vdc supply. This performance is still considerably below that of the more powerful CAF.

Performance Evaluations

Clinical intelligibility tests on a prototype of the Zeta self-adaptive filter coupled with a Rexton 25 PP body hearing aid were performed in 1984 at the

[1] The Zeta noise blocker is manufactured by Intellitech, Inc.

Table 1.
Intelligibility Tests Data for Four Groups of Subjects (11). Monosyllabic word intelligibility is shown averaged over 5 subjects in each group.

		600/800 Hz		1700/2400 Hz		Cafeteria		Babble		White Noise	
	In Quiet	NF	F	NF	F	NF	F	NF	F	NF	F
Group 1 Normals	97.6	44.2	73.6	40.0	58.4	38.4	68.4	36.8	56.0	39.2	44.0
Group 2 mild/ moderate	94.4	36.4	80.4	38.0	54.0	32.6	69.2	38.4	63.6	40.0	52.4
Group 3 moderate/ severe	77.6	37.6	64.0	40.0	48.0	36.0	69.6	37.6	46.0	39.2	38.4
Group 4 sloping high-freq. loss	74.8	40.8	53.2	40.8	54.0	42.8	46.8	42.4	52.8	34.0	48.0

Key: NF-nonfiltered, F-filtered

Siegal Ear Institute of the Michael Reese Medical Center, Chicago IL. (11). These tests were performed on 20 subjects, 5 with normal hearing, 5 with a mild-to-moderate sensorineural hearing loss, 5 with a moderate-to-severe sensorineural loss, and 5 with a steeply sloping high-frequency sensorineural loss.

Live voice readings of Northwestern University Test 6 (NU 6) were used with 5 noise conditions: 1) 600–800 Hz band-filtered noise (48 dB/octave); 2) 1700–2400 Hz band-filtered noise; 3) cafeteria noise; 4) six-speaker babbles, (3 male, 3 female); and, 5) white noise. Speech level was 68 dB SPL and noise level was varied by an up-down adaptive procedure to degrade listener's discrimination score to within the 30 to 50 percent range. Tests results are shown in **Table 1**.

It should be noted that the filtered scores should not exceed the score for the quiet condition since the filter cannot yield intelligibility scores (in the presence of noise) that are above those achieved for the same speech without any noise. For example, a 69.6 percent score in column "F" of **Table 1** for Group 3 with cafeteria noise implies a restoration of 89.7% of the maximum ("in quiet") score. For this group and noise condition without the filter (column NF), the average score was 36.0, or 46.4 percent of the maximum ("in quiet") score. The filter thus facilitated regaining of 89.7% − 46.4% = 43.3% of the maximum ("in quiet") score. The maximum possible score is 100 percent and the maximum attainable improvement is 100 − 46.4 = 53.6%. Hence, the percentage of maximum attainable score improvement provided by the filter was 43.3/53.6 namely, 80.8 percent under this noise condition. The results of the various noises for the 4 groups in terms of improvement as a percentage of maximal-attainable improvement are shown in **Table 2**.

Table 2.
Intelligibility improvement as a percent of maximum attainable improvement (11). The intelligibility improvement is equal to 100 · (F-NF)/(Quiet-NF), where the NF, F and Quiet conditions are as shown in **Table 1**.

	600/ 800Hz	1700/ 2400Hz	Cafeteria	Babble	White Noise
Group 1	55.1	31.9	50.7	31.6	8.3
Group 2	75.7	28.4	55.6	45.0	22.8
Group 3	66.0	21.3	80.8	21.0	−2.1
Group 4	36.5	38.8	12.5	32.1	34.3

Results of tests on the Zeta self-adaptive filter in actual chip-form when factory assembled in a commercial BTE hearing aid[2] are reported in Wollinsky (16) and reproduced in **Table 3**. These data are for

[2] Maico SP 345 (high gain).

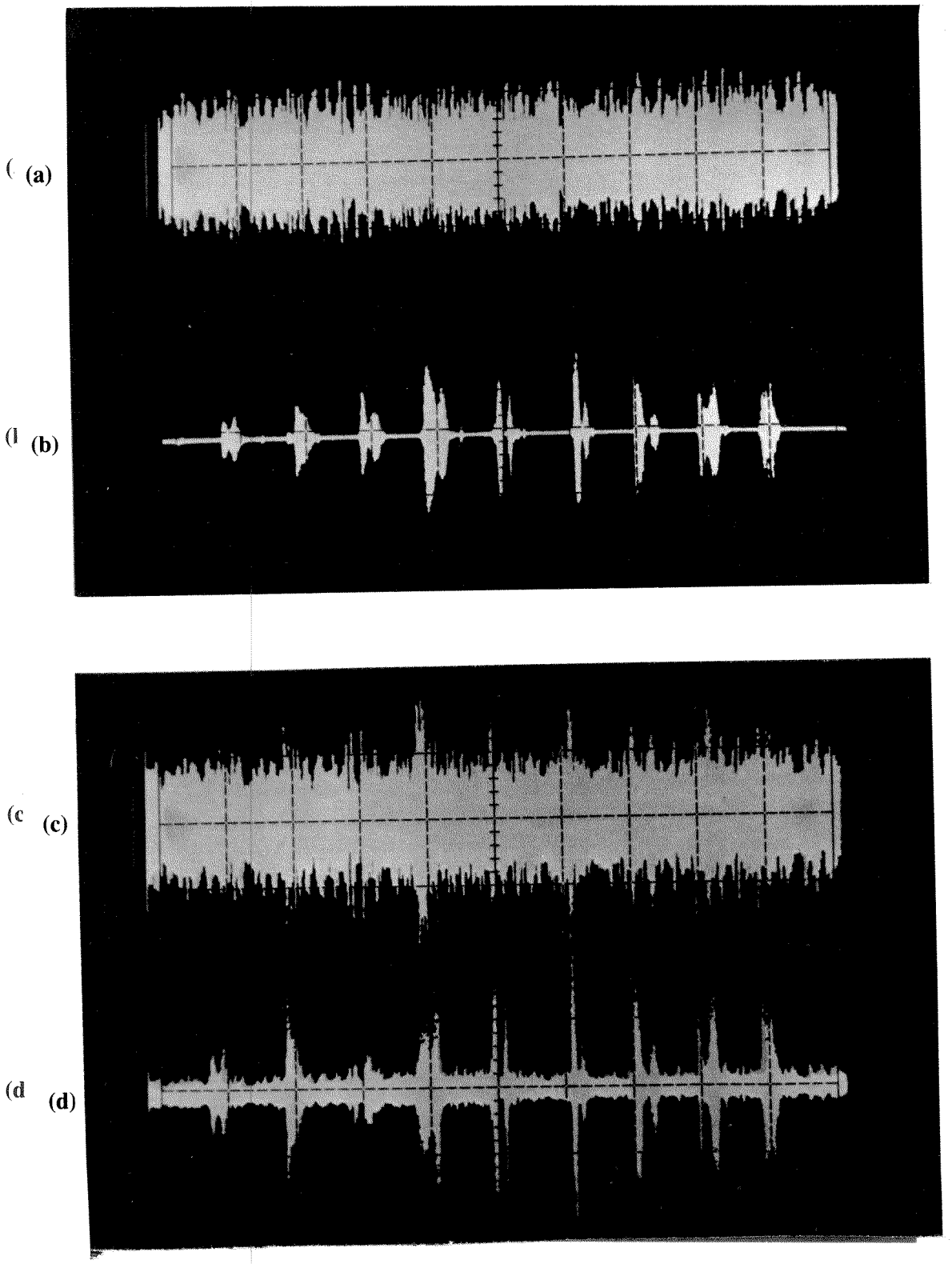

Figure 2.
Scope traces of filter's performance. Trace (a) shows 400–600 Hz Noise = n(t) of **Figure 1**. Trace (b) shows speech (counting from 2 to 10) = $\hat{s}(t)$ of **Figure 1**. Trace (c) shows speech plus noise at filter's input = y(t) of **Figure 1**. Trace (d) shows speech filtered by the Zeta self-adaptive filter = $\hat{s}(t)$ of **Figure 1**. All traces were recorded simultaneously in real time.

Table 3.
Intelligibility-improvement as a percent of maximum attainable improvement. Data are for a BTE hearing aid (16).

18 patient-average	Ave. Score In Quiet	Percent Intelligibility Improvement 600/800Hz	2700/3500Hz	Cafeteria	Babble
Moderate/Severe loss	66.1	82.0	43.0	55.0	23.0

18 hearing-impaired patients all in the moderate/severe loss group, using the same Northwestern University Monosyllabic (NU6) test tapes as for the data of **Tables 1 and 2**.

When comparing the scores of **Table 3** with those of **Table 2**, we note the Zeta noise-blocker chip used in the 1986 tests (16) was an improved version of the prototype used in the 1984 tests (11). As a result, slightly better performance was obtained in the 1986 study. Note also that the later study (16) used a BTE hearing aid while the earlier study (11) used a body aid. Further, in both tests, improvements for low-frequency (600–800 Hz) noises were higher than for high-frequency noises 1700–2400 or 2700–3500 Hz. This is to be expected because high-frequency noises are less likely to mask speech than low-frequency noises.

Test results performed by R. LaRose of Hinsdale, Illinois on the same prototype of the filter used by Stein and Dempsey-Hart (11) are given in **Table 4**. Intelligibility scores at various SNR ranging from −20 to +5 dB are shown. These data indicate that the filter performs well on normal-hearing persons at −20 dB SNR, in contrast to the methods discussed in the Introduction and Background section of this paper.

The extended (CAF) version of the Zeta self-adaptive filter is compared in terms of intelligibility scores and of SNR improvement with the single-chip version of the filter. The data are summarized in **Tables 5 and 6**, respectively. These data indicate an average improvement in performance of the extended filter over the hearing aid version by 35 to 40 percent.

The filter's performance is illustrated graphically in **Figure 2** in terms of the signal's time behavior. The top trace shown is the speech signal s(t), the second trace shows the noise n(t), the third trace shows speech and noise $y(t) = s(t) + n(t)$, and the lowest trace shows the output of the filter $\hat{s}(t)$. The traces are for exactly the same data and filtering run.

Table 4.
Intelligibility Scores at Various SNR.

Patient	Speech dB (HL)	Noise dB (HL)	SNR dB	Noise Type Hz	Score Non-Filtered	Filtered	Word List
Mrs. N.	65	60	+5	600/800	16 %	60 %	W-22
Normal	50	70	−20	600/800	24 %	80 %	NU-6
Mr. R.	60	65	−5	600/800	4 %	32 %	NU-6
Mr. N.	65	65	0	Babble	28 %	60 %	NU-6
Mrs. W.*	67.5	62.5	−5	Babble	0 %	24.5%	CID Sentences
Normal	50	70	−20	Babble	20 %	52 %	NU-6
Mr. N.	60	60	0	Cafeteria	36 %	84 %	W-22
Mrs. O.	57	67	−10	Cafeteria	0 %	64 %	W-22
Mrs. O.	57	62	−5	Cafeteria	64 %	94 %	CID Sentences
Mrs. O.	57	67	−10	Cafeteria	0 %	64 %	W-22
Mr. N.	60	65	−5	White Noise	0 %	16 %	W-22
Mrs. W.*	65	62.5	+2.5	White Noise	5 %	66 %	CID Sentences
Mrs. W.*	50	60	−10	2.4/3.0 KHz	45 %	92 %	CID Sentences

Mr. N.—Moderate to severe hearing loss 30–65 dB
Mr. R.—Mild loss 15–25 db, discrimination difficulties
Mrs. O.—Severe loss 55–75 db, low discrimination
*Mrs. W.—Mrs. W. scored nearly 0 at no noise with word lists, hence tested with sentence material

Table 5.
Intelligibility test scores: NU-6 monosyllabic word list.

Type of Noise	CAF 2		Zeta (1)	
	Unfil*	Fil**	Unfil	Fil
250–400 Hz	28%	72%	44.2%	73.6%(2)
1200–1700 Hz	36%	48%	52 %	64 %(3)
	16%	36%	72 %	72 %(4)
2400–3000 Hz	68%	74%	92 %	100 %(4)
Babble	26%	44%	36.8%	56 %(5)
Cafeteria	32%	90%	38.4%	68.4%

(% of fully correct word recognition)
*Unfiltered **Filtered
(1) For people with normal hearing, results from [11]
(2) Test for 600–800 Hz noise
(3) Tests by R. LaRose
Noise of 1200–1400 Hz
(4) Test by R. LaRose
(5) Test for 1700–2400 Hz

Table 6.
Improvement in signal-to-noise ratio (in dB) via the various filter designs.

	Improvement via Filter in dB			
Filter type	CAF		Zeta	
Type of Noise	Low Noise	High Noise	Low Noise	High Noise
250–400 Hz	+20.6	+25.5	+15.1	
2400–3000 Hz	+25.5		+10.35	
Babble	+ 8.4	+ 6.7	+ 6.3	+ 2.6
Cafeteria	+12.2	+16.2	+11.6	+12.7

* CAF 1 and CAF 2 refer to a time domain and frequency domain communication filter

SUMMARY

The Zeta self-adaptive filter described above and its CAF extension incorporate adaptive filtering and parameter identification theory with artificial-intelligence algorithms to result in a real time single-microphone adaptive filter. The chip is small enough to fit into standard BTE and ITE hearing aids. Because of its low-voltage CMOS realization, the Zeta self-adaptive filter operates with standard hearing aid batteries down to 1.1 Vdc, and is now available on several commercially produced hearing aids.

The performance of the Zeta self-adaptive filter and the related CAF self-adaptive filter shows significant improvement in intelligibility for a wide range of environmental noises. Significant improvement was also obtained at low SNR of −20 dB and less. The filter's performance in these respects has been tested both with hearing-impaired persons and with persons with normal hearing to determine its applicability beyond the hearing-impaired population.

REFERENCES

1. **Boll SF:** Suppression of accoustic noise in speech using spectral subtraction. *IEEE Trans Acoust Speech & Signal Proc* ASSP–29: 113–120, 1979.
2. **Christiansen RW, Chabries DM, and Anderson D:** Noise reduction in speech using a modifed LMS adaptive predictive filter. *Proc IEEE 7th Conf of the Eng in Med & Biol Soc,* 1985.
3. **Graupe D:** *Time Series Analysis, Identification and Adaptive Filtering.* Malabar, FL: Krieger Publishing Co., 1984.
4. **Graupe D, Grosspietsch JK, and Basseas SP:** A self adaptive filter of environmental noises from speech and its evaluation. *Proc IEEE CDC Conf,* 1985.
5. **Graupe D, Grosspietsch JK, and Basseas SP:** Self-adaptive filtering of environmental noises from speech. *Proc AIAA/IEEE 6th Avionics Sys Conf,* Baltimore, MD, 1984.
6. **Lentt H, Neuman A, Mills R, and Schwander T:** A digital master hearing aid. J Rehabil Res Dev 23(1): 79–87, 1986.
7. **Lim J S:** Enhancement of speech degraded adaptive noise: Overview. In *Speech Enhancement,* JS Lim (ed). Englewood Cliffs, NJ: Prentice Hall, 3–6, 1983.
8. **Lim JS:** Signal processing for speech enhancement. In *The Vanderbilt Hearing Aid Report,* G A Studebaker and FH Bess (eds). Upper Derby, PA: Monographs in Contemporary Audiology, 124–129, 1982.
9. **Lim JS and Oppenheim AV:** Enhancement and bandwidth compression of noisy speech. *Proc IEEE* 1586–1604, Dec. 1984.
10. **Sambur MR:** Adaptive noise cancelling for speech signals. *IEEE Trans Acoust Speech & Sign Proc* ASSP-26:419–423, 1979.

11. **Stein L and Dempsey-Hart D:** Listener-assessed intelligibility of a hearing aid self-adaptive noise filter. *Ear and Hearing* 5(4):199–204, 1984.
12. **Surr RK, Schuchman GI, and Montgomery AA:** Factors influencing use of hearing aids. *Hearing Instruments* 30:19, 1979.
14. **Weiss MR and Aschkenasy E:** Wideband speech enhancement. *RADC-TR-81–53, Final Tech Rep,* May 1981.
15. **Widrow B et al.:** Adaptive noise cancelling principles and applications. *Proc IEEE* 1692–1716, Dec. 1975.
16. **Wollinsky S:** Clinical assessment of a self adaptive noise filtering system. *The Hearing Journal* 29–32, Oct. 1986.