

07-20-00

A/PROV

PTO/SB/16 (6-95)
Approved for use through 04/11/98. OMB0651-0037
Patent and Trademark Office; U. S. DEPARTMENT OF COMMERCE

07/19/00
JCS68 U.S. PTO

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR § 1.53(c)

Express Mail label number EL473992223US Date of Deposit July 19, 2000
I hereby certify that this paper or fee is being deposited with the United States Postal Service
"Express Mail Post Office to Addressee" service under 37 CFR § 1.10
on the date indicated above and is addressed to the Commissioner for Patents, Washington, DC 20231.

Cindy Baglietto
Name of person signing

Cindy Baglietto
Signature

JCS41 U.S. PTO
60/219297
07/19/00

Docket Number	20628-701	Type a plus sign (+) inside this box →	+
---------------	-----------	---	---

INVENTOR(s)/APPLICANT(s)			
LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Burnett	Greg		San Francisco, California
TITLE OF THE INVENTION (280 characters max)			
METHOD AND APPARATUS FOR NOISE REMOVAL			
CORRESPONDENCE ADDRESS			
WILSON SONSINI GOODRICH & ROSATI 650 Page Mill Road Palo Alto, California 94304-1050 Telephone: (650) 493-9300 Facsimile: (650) 493-6811			
ENCLOSED APPLICATION PARTS (check all that apply)			
☒ Specification	Number of Pages <u>14</u>	☐ Small Entity Statement	
☐ Drawing(s)	Number of Sheets <u> </u>	☐ Other (specify) <u> </u>	
METHOD OF PAYMENT (check one)			
☐ A check or money order is enclosed to cover the Provisional filing fees.	PROVISIONAL FILING FEE AMOUNT (\$)	\$150.00	
☒ The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: <u>23-2415</u>			

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.
☐ Yes, the same of the U.S. Government agency and the Government contract numbers are:

Respectfully submitted,

SIGNATURE Barbara B. Courtney Date: 7-19-00

TYPED or PRINTED NAME Barbara B. Courtney REGISTRATION NO. 42,442
(if appropriate)

☐ Additional inventors are being named on separately numbered sheets attached hereto.

PROVISIONAL APPLICATION FILING ONLY

C:\NrPortb\PALib1\CB9\1208923_1.DOC

LGE EXHIBIT NO. 1009

Amazon v. Jawbone
U.S. Patent 8,019,091
Amazon Ex. 1009

00215297-071500

This is to describe the noise removal algorithm that I devised that takes advantage of our knowledge of when speech is occurring. It is simple, robust, and should work for any type of noise regardless of spectral content, duration, or stationarity. I call it the "Pathfinder" algorithm, as it uses the knowledge of when speech occurs to determine the transfer functions (paths) between two microphones.

Overview

One of the most common acoustic-only (as opposed to our algorithm which uses both acoustic and other information) adaptive noise removal algorithms is shown in Figure 1. It uses the input from two microphones and the normalized least-mean-squares (NLMS) method to adaptively remove noise. it works very well on sinusoidal inputs, even ones with 50 sinusoids and close to full bandwidth spectrum. it updates its noise parameters during unvoiced periods, thus assuring that it is only training itself on noise and not voiced information (although it may occasionally pick up unvoiced information if areas within ~500 msec of voiced speech are used for training). However, it does not work well on random noise or non-stationary noise.

The approach I took was to re-examine the assumptions behind this widely used noise removal technique. One of them is that the time that the signal is being received, that is the time that the signal $s(n) > 0$, is not known. For acoustic-only applications, this is certainly true. However, we

BUSINESS SENSITIVE AND CONFIDENTIAL

don't operate under this restriction. Using the energy of the GEMS signal, we know precisely when voiced speech is occurring. The question is then: What changes in the algorithm will occur if the times when $s(n)$ is nonzero are known?

If Figure 1, the acoustic information coming into Microphone 1 is denoted by $m_1(n)$. The information coming into Microphone 2 is similarly labeled $m_2(n)$. In the z (digital frequency) domain, we can represent them as $M_1(z)$ and $M_2(z)$. Then

$$\begin{aligned} M_1(z) &= S(z) + N_1(z) \\ M_2(z) &= N(z) + S_2(z) \end{aligned}$$

with

$$\begin{aligned} N_1(z) &= N(z)H_1(z) \\ S_2(z) &= S(z)H_2(z) \end{aligned}$$

so that

$$\begin{aligned} M_1(z) &= S(z) + N(z)H_1(z) \\ M_2(z) &= N(z) + S(z)H_2(z) \end{aligned} \tag{1}$$

This is the general case for all two microphone systems. There is always going to be some leakage of noise into Mic 1, and some leakage of signal into Mic 2. Equation 1 has four unknowns and only two relationships and cannot be solved explicitly.

However, perhaps there is some way to solve for some of the unknowns in Equation 1 by other means. Let's examine the case where the signal is not being generated – that is, where the GEMS signal indicates voicing is not occurring. In this case, $s(n) = S(z) = 0$, and Equation 1 reduces to

$$\begin{aligned} M_{1n}(z) &= N(z)H_1(z) \\ M_{2n}(z) &= N(z) \end{aligned}$$

where the n subscript on the M variables indicate that only noise is being received. This leads to

BUSINESS SENSITIVE AND CONFIDENTIAL

$$\begin{aligned} M_{1n}(z) &= M_{2n}(z)H_1(z) \\ H_1(z) &= \frac{M_{1n}(z)}{M_{2n}(z)} \end{aligned} \quad (2)$$

$H_1(z)$ can be calculated using any of the available system identification algorithms and the microphone outputs when only noise is being received. The calculation can be done adaptively, although if the relative position of the two microphones is held fairly constant $H_1(z)$ should be fairly constant as well. If done adaptively, the update would probably not need to be that often, perhaps on the order of once a second. The interesting thing is that the whiter the input, the better the transfer function calculation. That is, white noise, the most difficult type of noise to remove for acoustic-only systems, would actually result in the best performance for this technique.

So now we have solved for one of the unknowns in Equation 1. We can solve for another, $H_2(z)$, by using the amplitude of the GEMS or similar device along with the amplitude of the two microphones. When the GEMS indicates voicing, but the recent (less than 1 second) history of the microphones indicate low levels of noise, we can assume that $n(s) = N(z) \sim 0$. Then Equation 1 reduces to

$$\begin{aligned} M_{1s}(z) &= S(z) \\ M_{2s}(z) &= S(z)H_2(z) \end{aligned}$$

which in turn leads to

$$\begin{aligned} M_{2s}(z) &= M_{1s}(z)H_2(z) \\ H_2(z) &= \frac{M_{2s}(z)}{M_{1s}(z)} \end{aligned}$$

which is the inverse of the $H_1(z)$ calculation, but remember that we are using different inputs.

After we have calculated $H_1(z)$ and $H_2(z)$ above, we can use them to remove the noise from the signal. If we rewrite Equation 1 as

BUSINESS SENSITIVE AND CONFIDENTIAL

$$\begin{aligned}
S(z) &= M_1(z) - N(z)H_1(z) \\
N(z) &= M_2(z) - S(z)H_2(z) \\
S(z) &= M_1(z) - [M_2(z) - S(z)H_2(z)]H_1(z) \text{ ,} \\
S(z)[1 - H_2(z)H_1(z)] &= M_1(z) - M_2(z)H_1(z)
\end{aligned}$$

we can solve for $S(z)$:

$$S(z) = \frac{M_1(z) - M_2(z)H_1(z)}{1 - H_2(z)H_1(z)} \quad (3)$$

Summary

Given that we can determine the two cases when there is no noise and when there is no signal and can therefore calculate $H_1(z)$ and $H_2(z)$, the original signal may be recovered by Equation 3. Since both H_1 and H_2 should not change very rapidly, once determined we could just calculate

$$H_3(z) = \frac{1}{1 - H_2(z)H_1(z)}$$

and apply it until the next update of H_1 and H_2 occurs. We might not even need $H_3(z)$, it would just apply a magnitude and phase change to the audio, and that might not help intelligibility.

This would simplify the computations and result in a faster response.

The processing above would have to take place on finite windows of information, possibly 30 milliseconds or so in length or a certain number of glottal cycles. We could also try longer windows, but we don't want to delay the cleaned audio with respect to the original any more than necessary.

It is interesting to note that we are not making any approximations or assumptions regarding the noise or the signal. If we are able to calculate H_1 and H_2 sufficiently accurately, the noise will be completely removed regardless of the noise and signal characteristics.

BUSINESS SENSITIVE AND CONFIDENTIAL

1. This technique will allow the noise removal of any type of noise in the presence of any type of signal as long as the microphones are operating linearly and simply adding the signals – i.e. the signals are additive. If the microphones saturate, this algorithm will still work to a certain degree but will probably not remove the noise completely.
2. This technique does not depend on the location or type of microphones, nor does it require that the microphones be matched to each other. These details may be compensated for in the calculation of H_1 and H_2 .
3. This technique is not affected by aging of the microphones, as H_1 and H_2 can be recalculated whenever it is convenient. It is estimated that the calculation of H_1 and H_2 will take on the order of 10-100 milliseconds.
4. This technique does require the use of a “voicing device”, a device that can determine when the user is voicing (using the vocal folds to produce speech). This device would include but is not limited to radio frequency devices (such as the GEMS), electroglottographs (EGG), ultrasound devices, acoustic throat microphones, and airflow detectors.
5. The calculation of the transfer functions H_1 and H_2 can be accomplished by the use of any techniques used by those skilled in the art, including but not limited to adaptive techniques, recursive techniques, and simpler ones such as Matlab’s AR and ARX.
6. The calculation of the transfer functions need not be done constantly, as they should change slowly or not at all. The calculation of H_1 can be accomplished when the voicing device indicates that no voicing is occurring or has occurred within a set time. H_2 may be calculated when the recent history of the microphone signals indicate the absence of background noise and the voicing device indicates voicing is occurring.

BUSINESS SENSITIVE AND CONFIDENTIAL

006720 26267209
60249297 071900

multiple noise sources and reflective paths caused by the environment of the user (the multipath problem).

Overview

Several different circumstances can occur in the user's environment. In the first memo, we examined only the relatively simple case of a single noise source and a single signal source with only direct paths from the sources to the microphones. This situation is shown in Figure 1. We found that by determining when the signal was occurring, we could calculate $H_1(z)$ and $H_2(z)$ and remove all of the noise in this situation. The signal was given by

$$S(z) = \frac{M_1(z) - M_2(z)H_1(z)}{1 - H_2(z)H_1(z)}$$

this algorithm works quite well, using a normalized LMS algorithm to calculate H_1 and H_2 given the GEMS information on voicing. However, it only works on simulated data, not actual recorded data. The real world is just not as simple as that portrayed in Figure 1.

I will now generalize the Pathfinder algorithm to include many situations that are more realistic. The first is one in which there are multiple noise sources or equivalently, one noise source and

BUSINESS SENSITIVE AND CONFIDENTIAL

6

many reflections of the noise source. These situations are equivalent in that the microphones are unable to determine if a second noise input is a different noise source or just the original noise source after reflection from an interface. It is still assumed that there is only one path from the signal to the microphones.

Multiple noise sources, single signal with direct path

This case is illustrated in Figure 2. There are several noise sources illustrated, each with a transfer function (or path) to each microphone. The previously named path H_2 has been relabeled as H_0 , so that labeling noise source 2's path to Mic 1 is more convenient. The outputs of each microphone, when transformed to the z domain, are:

$$\begin{aligned} M_1(z) &= S(z) + N_1(z)H_1(z) + N_2(z)H_2(z) + \dots N_n(z)H_n(z) \\ M_2(z) &= S(z)H_0(z) + N_1(z)G_1(z) + N_2(z)G_2(z) + \dots N_n(z)G_n(z) \end{aligned} \quad \text{Eq. 1}$$

When there is no signal, as determined by the GEMS or other device, then (suppressing the z 's for clarity)

$$\begin{aligned} M_{1n} &= N_1H_1 + N_2H_2 + \dots N_nH_n \\ M_{2n} &= N_1G_1 + N_2G_2 + \dots N_nG_n \end{aligned} \quad (S = 0) \quad \text{Eq. 2}$$

We can define a new transfer function now, analogous to $H_1(z)$ in the previous memo:

$$\tilde{H}_1 = \frac{M_{1n}}{M_{2n}} = \frac{N_1H_1 + N_2H_2 + \dots N_nH_n}{N_1G_1 + N_2G_2 + \dots N_nG_n} \quad \text{Eq. 3}$$

Thus $\tilde{H}_1$ depends only on the noise sources and their respective transfer functions and can be calculated any time there is no signal being transmitted. Once again, the n subscripts on the microphone inputs denote only that noise is being detected, while an s subscript denotes that only signal is being received by the microphones.

If we now examine Equation 1 assuming that there is no noise, we get

$$\begin{aligned} M_{1s} &= S \\ M_{2s} &= SH_0 \end{aligned} \quad (N = 0)$$

So that H_0 can be solved for as before, using any available transfer function calculating algorithms desired. Mathematically

$$H_0 = \frac{M_{2s}}{M_{1s}}$$

BUSINESS SENSITIVE AND CONFIDENTIAL

If we now solve Equation 1, using $\tilde{H}_1$ above, we get

$$\tilde{H}_1 = \frac{M_1 - S}{M_2 - SH_0} \quad \text{Eq. 4}$$

Solving for S, we get

$$S = \frac{M_1 - M_2 \tilde{H}_1}{1 - H_0 \tilde{H}_1} \quad \text{Eq. 5}$$

which is the same as before, with H_0 taking the place of H_2 , and $\tilde{H}_1$ taking the place of H_1 . Thus the Pathfinder algorithm still is mathematically valid for any number of noise sources, including multiple echoes of noise sources. The only change in the algorithm written by Eric is that the estimates of $\tilde{H}_1$ contain both poles and zeros, whereas with a simpler FIR model there were only zeros. Still, if $\tilde{H}_0$ and $\tilde{H}_1$ can be estimated to a high enough accuracy, and the above assumption of only one path from the signal to the microphones holds, the noise may be removed completely.

Multiple noise sources, multiple signal sources

This case is the most general one possible and is illustrated in Figure 3. Here, we are allowing reflections of the signal to enter both microphones. This is the most general case, as reflections of the noise source into the microphones can be modeled accurately as simple additional noise sources. I have modified the names of the signal transfer functions, the direct path from the signal to Mic 2 has changed from $H_0(z)$ to $H_{00}(z)$, and the reflected paths to Microphones 1 and 2 are denoted by $H_{01}(z)$ and $H_{02}(z)$ respectively.

The input into the microphones now becomes

$$\begin{aligned} M_1(z) &= S(z) + S(z)H_{01}(z) + N_1(z)H_1(z) + N_2(z)H_2(z) + \dots N_n(z)H_n(z) \\ M_2(z) &= S(z)H_{00}(z) + S(z)H_{02}(z) + N_1(z)G_1(z) + N_2(z)G_2(z) + \dots N_n(z)G_n(z) \end{aligned} \quad \text{Eq. 6}$$

When the signal = 0, the inputs become (suppressing the z's again)

$$\begin{aligned} M_{1n} &= N_1H_1 + N_2H_2 + \dots N_nH_n \\ M_{2n} &= N_1G_1 + N_2G_2 + \dots N_nG_n \end{aligned} \quad (\mathbf{S} = \mathbf{0})$$

BUSINESS SENSITIVE AND CONFIDENTIAL

which is the same as Equation 2 above. Thus the calculation of $\tilde{H}_1$ in Equation 3 is unchanged, as expected. If we now examine the situation where there is no noise, Equation 6 reduces to

$$\begin{aligned} M_{1s} &= S + SH_{01} \\ M_{2s} &= SH_{00} + SH_{02} \end{aligned} \quad (\mathbf{N} = \mathbf{0})$$

This leads to the definition of $\tilde{H}_2$:

$$\tilde{H}_2 = \frac{M_{2s}}{M_{1s}} = \frac{H_{00} + H_{02}}{1 + H_{01}} \quad \text{Eq. 7}$$

If we again solve Equation 6 using the definition for $\tilde{H}_1$ (as in Equation 4), we get

$$\tilde{H}_1 = \frac{M_1 - S(1 + H_{01})}{M_2 - S(H_{00} + H_{02})} \quad \text{Eq. 8}$$

Some algebraic manipulation yields

$$\begin{aligned} S(1 + H_{01} - \tilde{H}_1(H_{00} + H_{02})) &= M_1 - M_2\tilde{H}_1 \\ S(1 + H_{01}) \left[1 - \tilde{H}_1 \frac{(H_{00} + H_{02})}{(1 + H_{01})} \right] &= M_1 - M_2\tilde{H}_1 \\ S(1 + H_{01}) [1 - \tilde{H}_1\tilde{H}_2] &= M_1 - M_2\tilde{H}_1 \end{aligned}$$

and finally

$$S(1 + H_{01}) = \frac{M_1 - M_2\tilde{H}_1}{1 - \tilde{H}_1\tilde{H}_2} \quad \text{Eq. 9}$$

Equation 9 is the same as equation 5, with the replacement of H_0 by $\tilde{H}_2$, and the addition of the $(1+H_{01})$ factor on the left side. This extra factor means that we cannot solve directly for S in this situation, but rather can only solve for the signal plus the addition of all of its echoes. This is not such a bad situation, as there are many conventional methods for dealing with echo suppression, and even if the echoes are not suppressed, it is unlikely that they will affect the comprehensibility of the speech to any meaningful extent. The more complex calculation of $\tilde{H}_2$ is needed to account for the signal echoes in Microphone 2, which act similar to noise sources. The effects of signal echoes into Mic 2 can therefore be compensated for, unlike those into Mic 1. This makes sense, as Mic 1 cannot determine if the signal it is receiving contains echoes and can therefore not remove them.

BUSINESS SENSITIVE AND CONFIDENTIAL

One thing that Equation 7 makes clear is that the addition of signal echoes makes $\tilde{H}_2$ dependent on the environment, not just the paths between the signal and the microphones. This means that $\tilde{H}_2$ will not be as constant as it is without signal echoes. Hopefully, though, this will be a small effect and not require frequent recalculation of $\tilde{H}_2$.

Conclusion

The Pathfinder algorithm has been shown to be viable under any environmental conditions. The type and amount of noise are inconsequential if a good estimate has been made of $\tilde{H}_1$ and $\tilde{H}_2$. If the user environment is such that echoes are present, they can be compensated for if coming from a noise source. If signal echoes are also present, they will affect the cleaned signal, but the effect should be negligible in most environments.

1. This technique will allow the noise removal of an arbitrary number of noise sources of any type in the presence of any type of signal. For example, this will work with low or high bandwidth noise, that is stationary or changing, that is short in duration or long, where there are 2 noise sources or 20.
2. This algorithm will function properly as long as the microphones are operating linearly and simply adding the signals – i.e. the signals are additive. If the microphones saturate, this algorithm will still work to a certain degree but will probably not remove the noise completely.
3. The amount of noise removal will depend on the accuracy of the calculation of $\tilde{H}_1$ and $\tilde{H}_2$. With accurate enough calculations, the noise will be completely removed.
4. This technique does not depend on the location or type of microphones, nor does it require that the microphones be matched to each other. These details may be compensated for in the calculation of $\tilde{H}_1$ and $\tilde{H}_2$.
5. This technique is not affected by aging of the microphones, as $\tilde{H}_1$ and $\tilde{H}_2$ can be recalculated whenever it is convenient. It is estimated that the calculation of $\tilde{H}_1$ and $\tilde{H}_2$ will take on the order of 10-100 milliseconds.

BUSINESS SENSITIVE AND CONFIDENTIAL

10

6. This technique does require the use of a “voicing device”, a device that can determine when the user is voicing (using the vocal folds to produce speech). This device would include but is not limited to radio frequency devices (such as the GEMS), electroglottographs (EGG), ultrasound devices, acoustic throat microphones, and airflow detectors.
7. The calculation of the transfer functions $\tilde{H}_1$ and $\tilde{H}_2$ can be accomplished by the use of any techniques used by those skilled in the art, including but not limited to adaptive techniques, recursive techniques, and simpler ones such as Matlab’s AR and ARX.
8. The calculation of the transfer functions need not be done constantly, as they should change slowly or not at all. The calculation of $\tilde{H}_1$ can be accomplished when the voicing device indicates that no voicing is occurring or has occurred within a set time. $\tilde{H}_2$ may be calculated when the recent history of the microphone signals indicate the absence of background noise and the voicing device indicates voicing is occurring. To increase the accuracy of the $\tilde{H}_1$ and $\tilde{H}_2$ calculations, the parameters may be averaged. A longer time constant should be used for $\tilde{H}_2$, as it will change quite slowly.

BUSINESS SENSITIVE AND CONFIDENTIAL

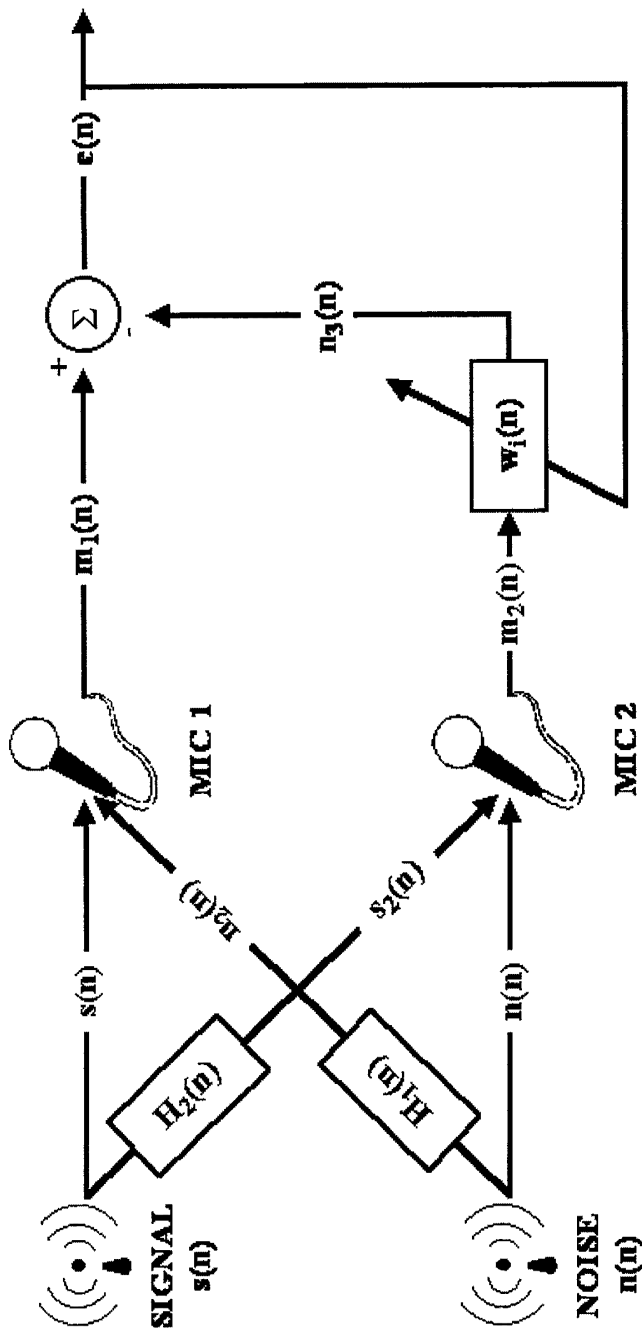
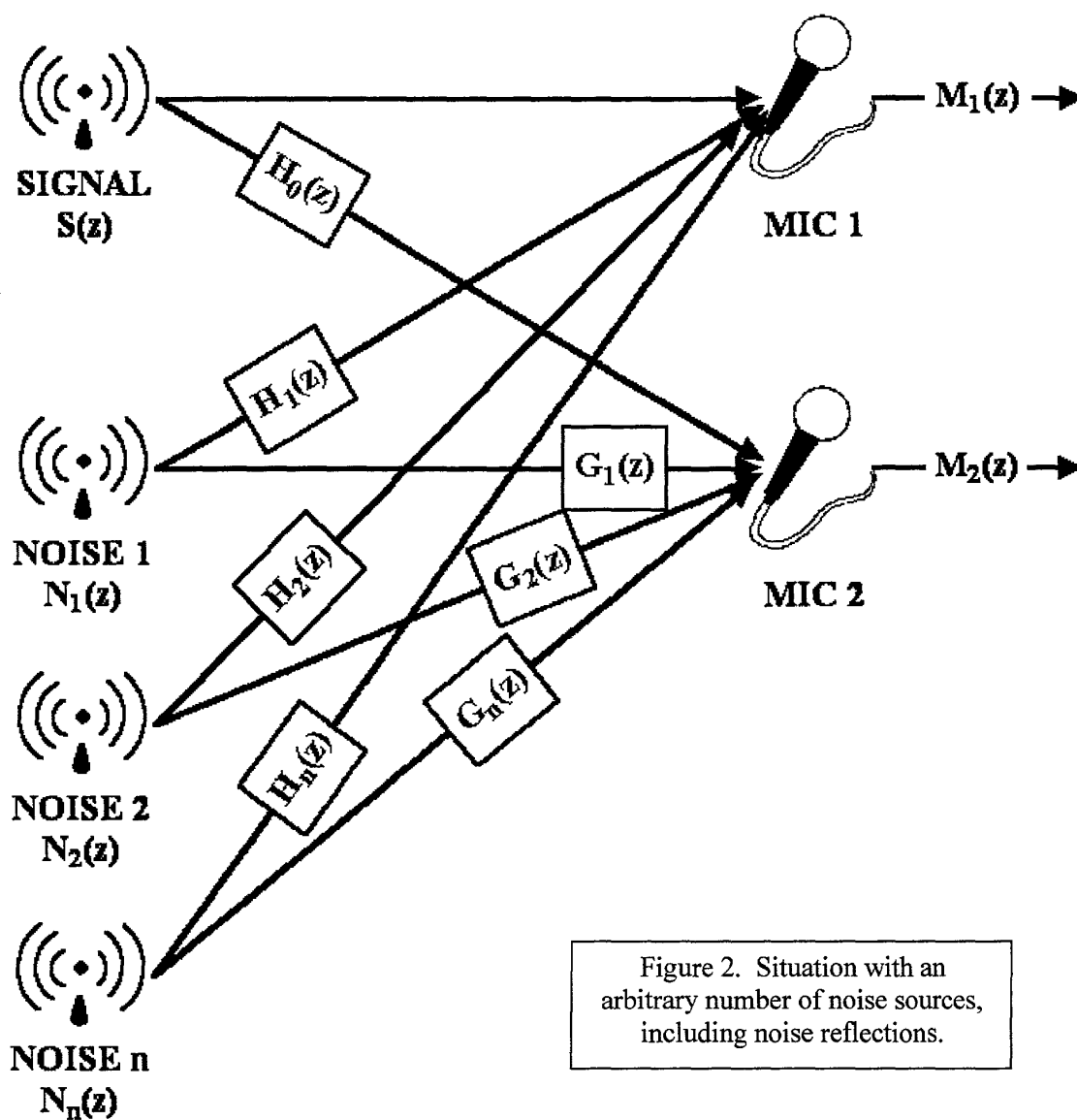
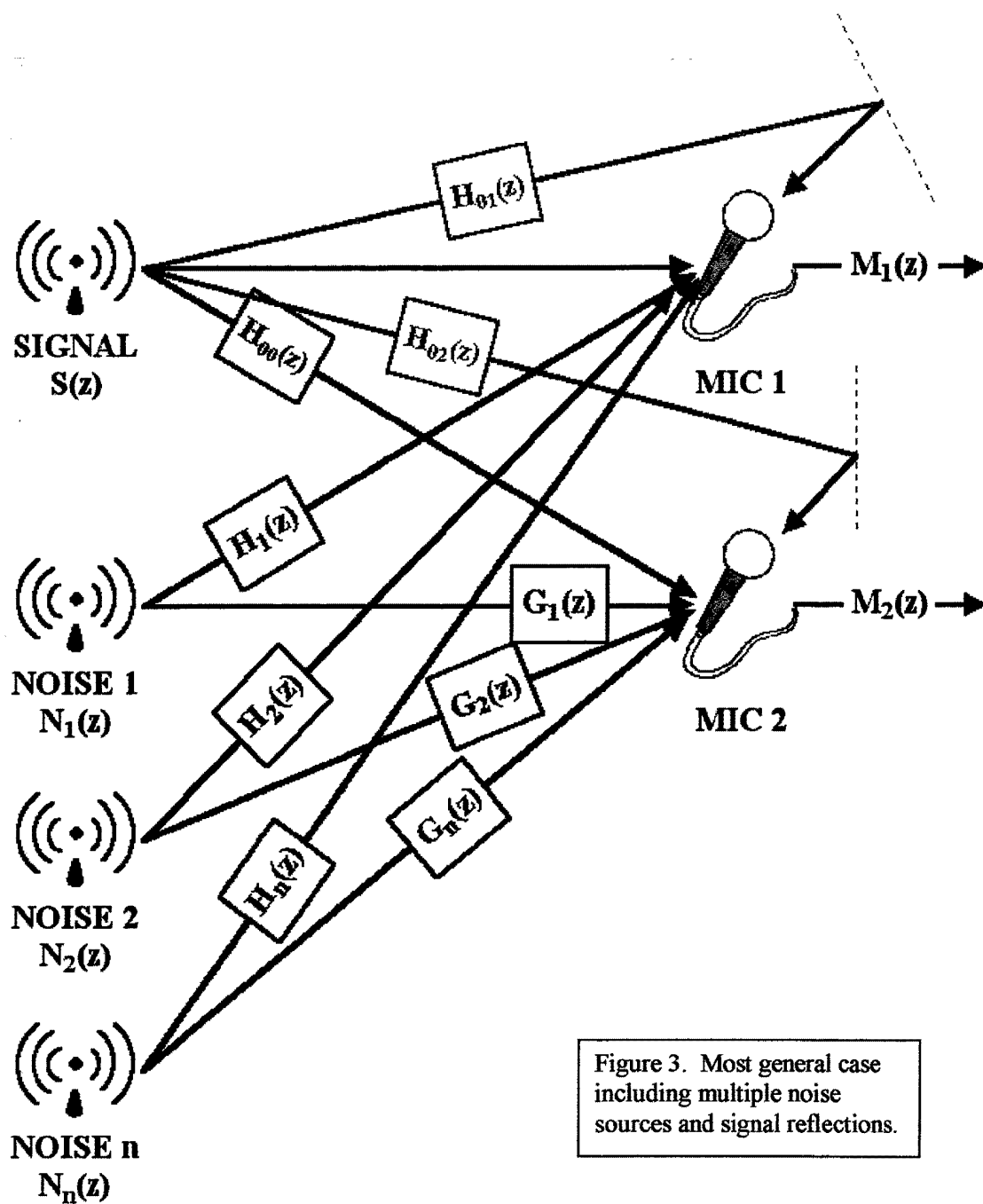


Figure 1. Setup for NLMS algorithm.



BUSINESS SENSITIVE AND CONFIDENTIAL



BUSINESS SENSITIVE AND CONFIDENTIAL