(11) Application No. AU 2004242561 A1

(19) AUSTRALIAN PATENT OFFICE

(54) Title
Modulation Depth Enhancement for Tone Perception

(51)[6] International Patent Classification(s)
G10L 021/02 H04R 025/00

(21) Application No: 2004242561 (22) Application Date: 2004.12.31

(30) Priority Data

(31) Number	(32) Date	(33) Country
2003907206	2003.12.31	AU

7

(43) Publication Date : 2005.07.14

(43) Publication Journal Date : 2005.07.14

(71) Applicant(s)
HearWorks Pty Ltd

(72) Inventor(s)
Van Hoesel, Richard; Vandali, Andrew

(74) Agent/Attorney
Watermark Patent & Trademark Attorneys, 290 Burwood Road, Hawthorn, VIC, 3122

2004242561 24 Mar 2005

ABSTRACT

A sound processing process is disclosed, with particular application to auditory prostheses. After input sound signals are processed into channels, an algorithm is applied to selectively increase the modulation depth of the envelope signals.

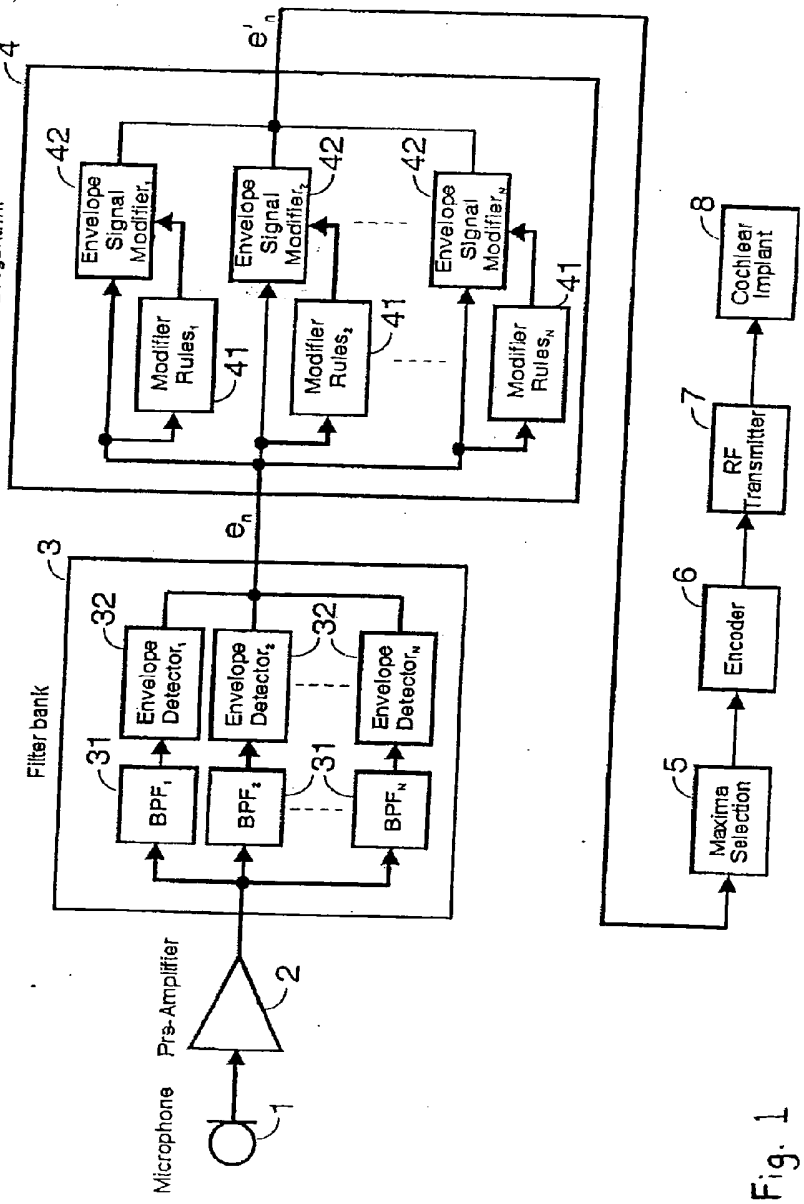
Microphone
1
Pre-Amplifier
2
Filter bank
3
BPF1
BPF2
BPFN
31
Envelope Detector1
Envelope Detector2
Envelope DetectorN
32
en
MDE Algorithm
4
Modifier Rules1
Modifier Rules2
Modifier RulesN
41
Envelope Signal Modifier1
Envelope Signal Modifier2
Envelope Signal ModifierN
42
e'n
Maxima Selection
5
Encoder
6
RF Transmitter
7
Cochlear Implant
8

Fig. 1

MODULATION DEPTH ENHANCEMENT FOR TONE PERCEPTION

FIELD OF THE INVENTION

This invention relates to the processing of signals derived from sound stimuli, particularly for the generation of stimuli in auditory prostheses, such as cochlear implants and hearing aids, and in other systems requiring sound processing or encoding. It further relates to software products and devices implementing such methods.

BACKGROUND OF THE INVENTION

Voice pitch information can play an important role in speech perception as it provides cues to linguistic features such as intonation (question – statement contrast) and word emphasis (Highnam, & Morris 1987; Nooteboom, 1997; Wells, Peppe, & Vance, 1995), and also to paralinguistic features such as speaker identification and the emotional state of the speaker (Abberton & Fourcin, 1978; Liberman, & Michaels, 1962) and segregation of concurrent speakers (Brokx, & Nooteboom, 1982). Most importantly, voice pitch information is crucial for perception of tonal languages, such as Mandarin and Cantonese, where a change in fundamental voicing frequency within the same phonemic segment causes a change in lexical meaning (Lee *et. al.*, 2002; Ciocca *et. al.*, 2002). Pitch information is also of important to the appreciation of music where the frequency of the fundamental and its harmonics governs the pitch of the signal (Moore 1995).

Various speech processing strategies have been developed for processing of sound signals for use in stimulating auditory prostheses, such as cochlear prostheses and hearing aids. The multi-peak strategy (Seligman, Dowell, & Blamey, 1992; Skinner *et. al.*, 1991) focused particularly on coding of aspects of speech, such as formants and the fundamental voicing frequency. For this strategy voice pitch information was predominantly coded by way of the electrical stimulation rate. Other strategies relied more on general channelization of the sound signal, such as the Spectral Maxima Sound Processor (SMSP) strategy, which is described in greater detail in Australian Patent No. 657959 and US Patent No. 5597390 by the present applicant. For this strategy voice pitch information (for a voicing frequency below approximately 200 Hz) is generally

2004242561 24 Mar 2005

coded in the envelope signals of each channel by amplitude modulation at a frequency equal to or related to the voicing frequency.

Experiments conducted with users of cochlear implant prostheses have indicated that the frequency of amplitude modulated electrical signals can be reliably detected when the modulation depth is sufficiently deep (McKay, McDermott, & Clark, 1994).

Channelization of the sound signal, as is done by most speech coding strategies today, often results in modulation depths within individual channels of less than 6 dB, even though the broadband sound signal has voicing frequency modulation of greater than 6 dB.

It is an object of the present invention to provide a sound processing strategy, and associated devices and software, to improve the user's perception of voice pitch and musical tone.

SUMMARY OF THE INVENTION

In a broad form, the present invention provides a sound processing strategy, of the type in which the sound signal is processed within defined frequency channels, wherein for at least some channels, when the modulation depth is less than a predetermined value, the modulation depth is increased.

Throughout the specification and claims, the term modulation depth in a channel means the ratio of the peak level to the trough level of the envelope signal in that channel over some finite time period.

According to one aspect, the present invention provides a sound processing process including at least the steps of:

(a) receiving a sound signal;

(b) processing said signal so as to produce a set of signals in spaced frequency channels; and

(c) performing further processing upon at least some of the set of signals;

wherein said process further includes the step of selectively increasing the modulation depth of the envelope signal for at least selected channels in response to a predetermined instruction set, prior to step (c).

From the prior studies of pitch perception, it appeared to the inventors that current channelization based speech processing strategies may not provide

2004242561 24 Mar 2005

adequate coding for identification of modulation frequency in the channel envelop signals. It was thus hypothesised that expansion of the envelope signal modulation depth in cases when it was shallow may provide improved identification of the modulation frequency and thus the voicing or musical pitch of the sound signal. The present invention is applicable to processing sound signals for auditory prostheses, including cochlear implants and hearing aids, as well as other applications where it may be desirable to improve the perception of voice pitch or musical tone.

If desired in particular applications, only some channels could be processed as defined above, although this is not presently preferred.

The modulation depth may be expanded by some constant function when it is below a given threshold, in a smoothly varying fashion, or by different functions at defined breakpoints. Alternative parameters could be adjusted, which have the effect of expanding the modulation depth.

BRIEF DESCRIPTION OF THE DRAWINGS

An illustrative embodiment of the present invention will now be described with reference to the accompanying drawings, in which:

Figures 1 and 2 are schematic representation of the signal processing applied to the sound signal in accordance with the present implementation;

Figure 3 depicts a typical input/output curve for the modulation depth;

Figures 4 and 5 are comparative electrodograms of sound signals to show the effect of the implementation;

Figures 6 and 7 are schematic representations of the signal processing applied by the MDE algorithm described in Appendix A – Approaches 2.A and 2.B respectively;

Figures 8 and 9 depict example envelope signals for a voiced passage of speech in a single channel and the subsequent modified envelop signals as processed by the MDE algorithm described in Appendix A.1 and A.2 respectively.

DETAILED DESCRIPTION

It will be appreciated that the present invention relates to an improvement which is applicable to a wide range of sound processing strategies for auditory prostheses, as well as other applications. Accordingly, the following

implementation is not to be taken as limitative of the scope or applicability of the present invention.

An implementation will be described with reference to a cochlear implant sound processing system. The precise system employed is not critical to the applicability of the present system. The present implementation will be described with reference to its use with the SMSP strategy (McDermott, & Vandali, 1991; McDermott, McKay, & Vandali, 1992), which is similar to the SPEAK strategy (Skinner *et. al.,* 1994; Whitford *et. al.,* 1995) and Advanced Combinational Encoder (ACE) strategy (Vandali *et. al.,* 2000). Note, however that it could equally be applied to other speech coding strategies such as the Continuous Interleaved Sampling (CIS) strategy (Wilson *et. al.,* 1991).

Referring to figures 1 and 2, as with the SMSP strategy, electrical signals corresponding to sound signals received via a microphone 1 and pre-amplifier 2 are processed by a bank of N parallel filters 3 tuned to adjacent frequencies (typically N = 16 for the conventional SMSP but in this implementation N can be varied and typically = 20). Each filter channel includes a band-pass filter 31 and an envelope detector 32 to provide an estimate of the narrow-band envelope signal in each channel. The band-pass filters are typically narrow (approximately 180 Hz) for apical (low-frequency) channels and increase in bandwidth (typically up to 1000 Hz or more) for more basal (higher frequency) channels. The envelope detectors, which effectively comprise a full-wave (quadrature) rectifier followed by a low-pass filter, typically pass fundamental (modulation) frequency information up to approximately 180 Hz to 400 Hz but for some implementations higher frequencies can be passed.

In this embodiment either a Fast Fourier Transform (FFT) or a Finite Impulse Response (FIR) filter bank (which uses complex coefficients) could be employed to implement the filter bank. Both implementations effectively perform the band-pass filtering, full-wave (quadrature) rectification and low-pass filtering. The FFT filter bank provides a fixed low-pass filter cut-off frequency (for –3 dB gain) of 180 Hz. The complex coefficient FIR provides a low-pass filter cut-off frequency equal to the (–3 dB) bandwidth of the band-pass filters. Basal (high frequency) channels can be as wide as 1000 Hz or more and thus an additional 2^{nd} order low-pass filter (with a cut-off frequency of 400 Hz) can optionally be

2004242561 24 Mar 2005

employed to remove any fine structure above the fundamental voicing frequency from the envelope signals. The advantage of employing the complex coefficient FIR over the FFT filter bank method is that higher voicing frequencies can be passed, provided that the band-pass filters are wider than 180Hz.

The filter bank is used to provide an estimate of the envelope signals in each channel at regular time intervals known as the analysis or update rate. The SMSP strategy conventionally employs a relatively low analysis rate of approximately 250 Hz, however in this implementation a much higher update rate of approximately 1200 to 1600 Hz is employed so that modulation frequencies of approximately 300 to 400 Hz can be adequately sampled. Such update rates are available with current commercial cochlear implant systems and speech coding strategies such as ACE. It is known from amplitude modulation identification experiments with users of cochlear implant prostheses that update/stimulation rates of at least four times the modulation frequency are required for adequate analysis/coding of the signal (McKay, McDermott, & Clark, 1994).

The outputs of the N-channel filter bank are modified by the Modulation Depth Enhancement (MDE) algorithm 4, as described below, prior to further processing by the speech coding strategy. The MDE algorithm operates on the narrow-band envelope signals in each filter bank channel independently. The envelope signals in each channel are analysed so as to estimate the modulation depth 412 (i.e. the ratio of peak-to-trough amplitude 411) over some finite time period (τ).

The estimated modulation depth in each channel (MD_n), where n refers to the channel number, is defined as shown in Eq. 1 below.

$$MD_n = P_n / T_n \qquad (1)$$

where P_n = maximum (Peak) level and T_n = minimum (Trough) level of the envelope signal in each channel over some finite time period and are determined using a sliding time window of duration (τ).

The duration of the sliding time window (τ) is typically 10 to 15 ms and is sufficiently long enough to analyse fundamental voicing frequencies as low as 100 Hz. For periodic voiced signals such as vowels, the maximum and minimum

levels will respectively follow peak and trough envelope signal levels relatively accurately provided that: the voicing period is shorter than the duration (τ) of the sliding time window; and that modulations in the signal at higher harmonic frequencies than the fundamental do not interfere with the modulation depth of the fundamental. For un-voiced signals, such as friction, which have no specific periodicity, the peak and trough levels (and thus the estimated modulation depth) can vary greatly from one peak-trough cycle to the next.

Because the modulation depth is estimated over some finite duration, rather than instantaneously, the estimate must be referenced from a time point corresponding to the middle of the time window. Thus a processing delay of $\tau/2$ is introduced for all processing following the modulation depth estimation.

The estimated modulation depth (and hence the envelope signal) for each channel is modified 42, according to some rules (or input/output function) 41, so as to effectively increase the modulation depth in cases when it is small (shallow).

It will be appreciated that many alternatives exist for expanding the modulation depth, and that the example in this implementation is only one alternative. Various alternative implementations of the MDE algorithm are provided in Appendix A. The algorithm described in the first approach from Appendix A – Approach 1.A is summarised below.

In this implementation, a power function is used to expand the modulation depth for cases when it is less than some knee point (typically 6 dB). The order of the power function is typically 2 or 3. For modulation depths greater than this knee point but less than some limit (typically 20 dB), a linear function is used to modify the modulation depth. For modulation depths above this limit point the modulation depth is unchanged.

One possible set of rules for modification of the modulation depth 413 are defined as follows:

(a) For modulation depths less than or equal to some Knee point (K_{MD}), which typically equals 2 (6 dB), the modified modulation depth (MD'_n) is increased using a power function where the Expansion Factor (X_{MD}), which is typically equal to 2 or 3, defines the order of the power function.

$$MD'_n = MD_n^{X_{MD}} \qquad \text{for} \qquad MD_n \leq K_{MD} \qquad (2)$$

(b) For modulation depths greater than the Knee point but less than some Limit point (L_{MD}), which typically equals 10 (20 dB), the modulation depth is still increased but a linear function is employed. The constants A and B are calculated such that boundary conditions are satisfied (i.e. no discontinuities) at the knee and limit points.

$$MD'_n = MD_n \times A + B \qquad (3)$$

for $K_{MD} < MD_n < L_{MD}$ and $K_{MD}^{X_{MD}} < L_{MD}$

where $A = (L_{MD} - K_{MD}^{X_{MD}}) / (L_{MD} - K_{MD})$ and $B = L_{MD} \times (1 - A)$

(c) For modulation depths above the Limit point the modulation depth is left unchanged.

$$MD'_n = MD_n \quad \text{for} \quad MD_n \geq L_{MD} \qquad (4)$$

Figure 3 depicts an input/output curve, plotted on a log-log dB scale, for the modulation depth using a Knee point of 2 (6 dB) a Limit point of 10 (20 dB) and an Expansion factor of 3.

The modified modulation depth MD'_n is used to adjust the trough T_n 414 level of the envelope signal such that the modified trough level T'_n is reduced by the ratio of the original modulation depth over the modified modulation depth.

$$T'_n = T_n \times MD_n / MD'_n = P_n / MD'_n \qquad (5)$$

However for points in time where the envelope signal is not at a trough, the envelope signal must be modified (e'_n) based on the required reduction to the trough level. A linear equation (Eq. 6) can be employed to modify the continuum of levels in the envelope signal 42. The use of a linear function preserves the shape of the envelope signal within each voicing period (or periodic cycle).

$$e'_n = e_n \times C_n + D_n \qquad (6)$$

where $C_n = (P_n - T'_n) / (P_n - T_n)$ and $D_n = P_n \times (1 - C_n)$

Solutions for C_n and D_n 415 (and thus MD'_n 413 and T'_n 414) are computed when either the peak or trough levels change. Solution of e'_n 42 is conducted for every time point in the envelope signal. Figure 8 displays an example unmodified e_n and modified e'_n envelope signal in one channel for a typical voiced passage of speech.

It will be appreciated that the parameters used represent only one possible strategy possible under the implementation described. For example, the inventors have trailed alternative parameters for the strategy. One form uses a knee point of 10dB, a limit point of 80 dB, and an expansion power of 7 (below the knee point). This provides a greater expansion of modulation depth. Another alternative provides more moderate expansion, with a knee point of 6dB, a limit of 40dB, and an expansion power of 4.

The modified envelope signals e'_n replaces the original envelope signals e_n derived from the filter bank and processing then continues as per the original speech coding strategy. For the SMSP strategy (or the SPEAK and ACE strategies) M of the N channels of e'_n having the largest amplitude at a given instance in time are selected 5 (typically M = 8 for this embodiment). The M selected channels are then used to generate M electrical stimuli 6 corresponding in stimulus intensity and electrode number to the amplitude and frequency of the M selected channels. These M stimuli are transmitted to the Cochlear implant 8 via a radio-frequency link 7 and are used to activate M corresponding electrode sites. The modulation depth enhancement may be applied to the channelised sound signal, and subsequent processing continue as per any selected processing strategy for the cochlear implant. This strategy is specific to this stage of processing, and hence is applicable to any strategy which employs channelization and subsequent processing (with modifications as may be dictated by the requirements of the selected strategy).

To illustrate the effect of the strategy on the coding of speech signals, stimulus output patterns, known as electrodograms (which are similar to spectrograms for acoustic signals), which plot stimulus intensity (plotted as log current level) for each electrode (channel) as a function of time, were recorded for the SMSP and MDE strategies and are shown in Figs. 4 & 5 respectively. The

speech token presented in these recordings was "lime" and was spoken by a female speaker having a fundamental voicing frequency of approximately 200 Hz. Note, the electrodogram for the MDE strategy depicts the response for the algorithm as described in Appendix A - Approach 2.B. The MDE Knee point was set to 6 dB, the Limit point to 20 dB and the Expansion factor to 3. The effect of the MDE strategy can be seen by comparing Figs. 4 and 5. For cases where the un-modified modulation depth is small or less than the Knee point (e.g. points A, B and C), the modified modulation depth is expanded by a factor of approximately 3 on a log scale. For cases where the unmodified modulation depth is above the knee point but below the Limit point (e.g. points D, E and F) the modulation depth is still expanded but by a factor less than 3 which approaches 1 as the modulation depth approaches the Limit point. For cases where the unmodified modulation depth is above the Limit point (e.g. points G, H and I), the modulation depth is unmodified. Note, for unvoiced or noisy segments of the signal (e.g. point J) the modulation depth is still modified.

In trials, the inventors have observed the best results in temporal pitch perception have been obtained when the technique described above is combined with a strategy to align temporal peaks across channels. A detailed description of this strategy is annexed as appendix C. In brief, this strategy is applied after the modulation depth has been expanded according to the present invention. The envelope for each channel is determined and temporal peaks identified. A timing offset is then selectively applied to each channel signal, so that the phase differences between the temporal peaks are reduced. These phase adjusted signals are then used as the basis for further processing.

Appendix B provides details on how the MDE algorithm might be implemented in a real-time DSP processing system.

The reader will appreciate that the present invention is of broad application, and that additions or modifications are readily possible within the broad inventive concept disclosed.

APPENDIX A: DESCRIPTION AND DERIVATION OF THE MDE ALGORITHM

Referring to Figures 1 and 2 for each channel the modulation depth (MD_n), where n refers to channel number, or by definition peak-to-trough ratio of the envelope signal in each channel (e_n) can be estimated 412 by dividing the

maximum (peak P_n) by the minimum (trough T_n) levels in e_n as determined 411 over some finite duration using a sliding time window (refer to Eq. A1.1). The duration of the sliding time window (τ) is typically 10 to 15 ms and is sufficiently long enough to analyse fundamental voicing frequencies as low as 100 Hz.

$$MD_n = P_n / T_n \qquad \text{(A1.1)}$$

where P_n = Maximum level of e_n over sliding time window of duration τ
and T_n = Minumum level of e_n over sliding time window of duration τ

Note, for periodic voiced signals such as vowels, the maximum and minimum levels will respectively follow peak and trough envelope signal levels relatively accurately provided that: the voicing period is shorter than the duration (τ) of the sliding time window; and that modulations in the signal at higher harmonics frequencies than the fundamental do not interfere with the modulation depth of the fundamental. For un-voiced signals, such as friction, which have no specific periodicity, the peak and trough levels (and thus the estimated modulation depth) can vary greatly from one peak-trough cycle to the next.

Because the modulation depth is estimated over some finite duration, rather than instantaneously, the estimate must be referenced from a time point corresponding to the middle of the time window. Thus a processing delay of $\tau/2$ is introduced for all processing following the modulation depth estimation. This time shift is to be assumed for the remainder of this description.

The estimated modulation depth (MD_n) in each channel could be modified (MD_n') according to some rules that effectively increase the modulation depth in cases when the modulation depth is small.

APPROACH 1

One possible set of rules 413 that could be used to implement this are described as follows:

(a) For modulation depths less than or equal to some Knee point (K_{MD}), the modified modulation depth is increased using a power function where the

Expansion Factor (X_{MD}), which is typically equal to 2 or 3, defines the order of the power function.

$$MD'_n = MD_n^{X_{MD}} \quad \text{for} \quad MD_n \leq K_{MD} \qquad \text{(A1.2)}$$

(b) For modulation depths greater than the Knee point but less than some Limit point (L_{MD}), the modulation depth is still increased but a linear function (refer to Eq. A1.3) is used to adjust the modulation depth. The constants A and B can be derived for the following boundary conditions: MD'_n equals $MD_n^{X_{MD}}$ at the knee point (i.e. when $MD_n = K_{MD}$) and MD_n is unchanged (i.e. $MD'_n = MD_n$) at the Limit point. Note, K_{MD} raised to the power of X_{MD} must be less than L_{MD}.

$$MD'_n = MD_n \times A + B \qquad \text{(A1.3)}$$

for $K_{MD} < MD_n < L_{MD}$ and $K_{MD}^{X_{MD}} < L_{MD}$

where $A = (L_{MD} - K_{MD}^{X_{MD}}) / (L_{MD} - K_{MD})$

and $B = L_{MD} \times (1 - A)$

(c) For modulation depths above the Limit point the modulation depth is unchanged.

$$MD'_n = MD_n \quad \text{for} \quad MD_n \geq L_{MD} \qquad \text{(A1.4)}$$

Figure 3 depicts an input/output curve, plotted on a log-log dB scale, for the modulation depth using a Knee point of 2 (6 dB) a Limit point of 10 (20 dB) and an Expansion factor of 3.

The envelope signals (e_n) are modified (e'_n) so as to achieve the desired modifications to the modulation depth. Recall that the modulation depth is equal to the peak-to-trough ratio of e_n calculated over some finite interval (τ). Thus to increase the modulation depth either the peak level could be increased, the trough level could be decreased, or some function of both increasing the peak and decreasing the trough could be carried out. In order to minimise loudness

changes when modifying the modulation depth, it might be desirable to keep the average level of the envelope signal constant. Thus both the peak and trough levels could be adjusted so as to preserve the average level. This approach would be recommended for non-cochlear implant prosthesis (such as hearing aids). However for cochlear implant prostheses, peaks of electrical stimulation contribute mostly to the perceived loudness of the signal and thus to minimise loudness changes, the peaks should be preserved and only the troughs of the envelope signals should be modified.

For cases when the envelope signal is at a trough (i.e. when $e_n = T_n$) the relation shown in Eq. (A1.5) can be used to determine the modified trough level (T'_n) which is inversely proportional to the ratio of the modified modulation depth over the original modulation depth 414.

$$T'_n = T_n \times MD_n / MD'_n = P_n / MD'_n \qquad \text{(A1.5)}$$

However for points in time where the envelope signal is not at a trough, the modified values for the envelope signal (e'_n) need to be calculated based on the required reduction to the trough level.

APPROACH 1A

A simple linear equation could be used to modify the continuum of levels in the envelope signal. The use of a linear function will preserve the shape of the envelope signal within each voicing period (cycle). This linear equation could be of the form shown in Eq. (A1.6). Constants C_n and D_n could be derived 415 such that: the envelope signal is unchanged (i.e. $e'_n = e_n$) when the envelope signal is at a peak; and the envelope signal is adjusted according to the desired modulation depth increase (i.e. $e'_n = e_n \times MD_n/MD'_n$) when the envelope signal is at a trough.

$$e'_n = e_n \times C_n + D_n \qquad \text{(A1.6)}$$

where $C_n = (P_n - T'_n) / (P_n - T_n)$ and $D_n = P_n \times (1 - C_n)$

Solutions for C_n and D_n 415 (and thus MD'_n 413) would only need to be sought when either the peak or trough levels change. Solution to e'_n 42 would be carried out for every time point in the envelope signal.

Alternate functions rather than a linear equation for modification of the continuum of levels in the envelope signal could be employed. For instance, it may be desirable to better preserve the peak level by using a 2nd or higher order equation that adjusts levels in the trough region (i.e. levels below the average or mid-point of the envelope signal) more than those in the peak region (i.e. levels above the mid-point of the envelope signal). This would ensure less change to the loudness of the peaks and thus less change to the overall loudness of the perceived signal after processing. It will however distort the shape of the envelope signal within each voicing period.

APPROACH 1B

Rather than using a 2nd or higher order equation an alternate approach might use a linear equation but change the boundary conditions such that only trough regions (i.e. levels below the mid-point of the envelope signal) are modified. The mid-point of the envelope signal could be defined as follows:

$$M_n = (P_n + T_n) / 2 \qquad \text{(A1.7)}$$

For cases when the envelope signal is above the mid-point no change to the signal would be applied. However for cases when the envelope signal is below the mid-point a linear equation could be employed to modify the signal such that mid-point levels are unchanged but levels at a trough are decreased by the desired increase to the modulation depth. The same linear equation as used above (i.e. Eq. A1.6) could be used but the constants C_n and D_n would be adjusted by making reference to the mid-point M_n.

$$e'_n = e_n \times C_n + D_n \qquad \text{for} \quad e_n < M_n \qquad \text{(A1.8)}$$

where $C_n = (M_n - T'_n) / (M_n - T_n)$ and $D_n = M_n \times (1 - C_n)$

Note, the above approach will preserve the shape of the signal when it is above the mid-point and then stretch the signal when it is below the mid-point. Figure 9 displays an example unmodified (e_n) and modified (e'_n) envelope signal for this approach. Note also that the schematic shown in Figure 2 for Approach 1.A also applies for Approach 1.B, however an extra calculation for the mid-point level (M_n) is required in 411 and calculation of C_n and D_n in 415 will be relative to the mid-point rather than the peak level.

APPROACH 2

APPROACH 2A

An alternative approach for implementation of the algorithm could adjust the envelope signal based on the "signal depth" rather than the total modulation depth. Referring to Figure 6, which replaces the processing shown in Figure 2, the "signal depth" (sd_n) 432 could represents the ratio of the peak level 431 to the envelope signal level at any time point and be defined as shown in Eq. (A2.1). The "signal depth" will equal the true modulation depth when the signal is at a trough and will equal unity when the signal is at a peak. For all levels between the peak and trough the "signal depth" will be inversely proportional to the signal level.

$$sd_n = P_n / e_n \qquad \text{(A2.1)}$$

The "signal depth" is calculated continuously and used to adjust the envelope signal level for all time points. Applying similar rules to those used in Approach 1 for modification of the modulation depth, and using the relation $e'_n = P_n / sd'_n$ we can establish equations for the modified envelope signal levels as a function of the "signal depth" 44.

(a) For "signal depths" less than or equal to the Knee point:

$$e'_n = P_n / sd_n^{X_{MD}} \qquad \text{(A2.2)}$$

for $sd_n \leq K_{MD}$

(b) For "signal depths" greater than the Knee point but less than the Limit point:

$$e'_n = P_n / (sd_n \times A + B) \qquad \text{(A2.3)}$$

for $K_{MD} < sd_n < L_{MD}$ and $K_{MD}^{X_{MD}} < L_{MD}$

where $A = (L_{MD} - K_{MD}^{X_{MD}}) / (L_{MD} - K_{MD})$

and $B = L_{MD} \times (1 - A)$

(c) For "signal depths: above the Limit point the envelope signal level is preserved.

$$e'_n = e_n \quad \text{for} \quad sd_n \geq L_{MD} \qquad \text{(A2.4)}$$

APPROACH 2.B

As in Approach 1.B the loudness of the processed signal might be better preserved by restricting signal modification to time points in which the envelope signal is less than the mid-point (M_n as defined in Eq. A1.7) between its peak and trough levels. In addition, computational time would be reduced as calculation of the "signal depth" is expensive as it requires a divide operation.

Modification of the envelope signal could simply be restricted to points in which the envelope signal is less than the mid-point (i.e. $e_n < M_n$). As pointed out in Approach 1.B this will introduce distortion of the envelope signal (i.e. a step change in the envelop signal level) at values around the mid-point. However, for cochlear implant prostheses this may not pose a big problem as it is unlikely that this sort of distortion is noticeable or destructive to the signal. For non-cochlear implant prostheses (such as hearing aids) this sort of distortion may be noticeable and should be avoided. In fact even Approach 2.A can introduce inter-period distortion that may be noticeable and thus Approach 1 is recommended for non-cochlear implant prostheses.

The distortion discussed above may be alleviated somewhat by re-defining the equation for the "signal depth", as shown in Eq. (A2.1), as a function of the mid-point, rather than the peak, of the envelope signal level.

$$sd_n = M_n / e_n \qquad \text{for} \quad e_n < M_n \qquad \text{(A2.5)}$$

However the "signal depth" sd_n now no longer equals the modulation depth when the signal is at a trough. Modifying Eq. (A2.5) so that the boundary conditions of: $sd_n = 1$ at the mid-point (i.e. for $e_n = M_n$); and $sd_n = MD_n$ at a trough (i.e. for $e_n = T_n$) are met we obtain:

$$sd_n = (2 \times M_n - e_n) / e_n \qquad \text{for} \quad e_n < M_n \qquad \text{(A2.6)}$$

Referring to Figure 7, which replaces the processing shown in Figure 6, similar modulation depth rules as used above in Eqs. (A2.2) to (A2.4) can be used to derive equations for the modified envelope signal (e'_n) 46 as a function of the "signal depth" 452 as defined in Eq. (A2.6) for all time point in e_n which are less than the mid-point M_n 451.

(a) For "signal depths" less than or equal to the Knee point:

$$e'_n = M_n / sd_n^{(X_{MD} - 1)} \qquad \text{(A2.7)}$$

for $sd_n \leq K_{MD}$ and $e_n < M_n$

(b) For "signal depths" greater than the Knee Point but less than the Limit point:

$$e'_n = M_n / (sd_n \times A + B) \qquad \text{(A2.8)}$$

for $K_{MD} < sd_n < L_{MD}$ and $K_{MD}^{X_{MD}} < L_{MD}$ and $e_n < M_n$

where $A = (L_{MD} - K_{MD}^{X_{MD}}) / (2 \times (L_{MD} - K_{MD}))$

and $B = L_{MD} \times A + (1 - L_{MD}) / 2$

(c) For "signal depths: above the Limit point the envelope signal level is preserved.

$$e'_n = e_n \qquad \text{for} \qquad sd_n \geq L_{MD} \qquad \text{(A2.9)}$$

Note however that the rules differ slightly from those in Approach 2.A because the "signal depth" is now relative to the mid-point, rather than the peak. In addition, for cases in which the modulation depth (or "signal depth") is small (i.e. less than the knee point), the modulation depth expansion factor will be less than X_{MD} (i.e. approximately $X_{MD} - 0.5$).

APPENDIX B: Conversion of MDE algorithm to a form suitable for real-time DSP processing

In converting the MDE algorithm to a form suitable for real-time DSP processing two main criteria need to taken into consideration. Firstly fixed-point DSP processing deals with numerical values less than or equal to 1.0. Thus when dealing with parameters in the algorithm such the modulation depth (i.e. the ratio of the peak-to-trough level which is a value that is always greater than or equal 1.0) the parameters must either be scaled such that they fall into a usable range below 1.0, or inverted (i.e. reciprocal) such that they will never be greater than 1.0. For the case of the modulation depth (and "signal depth") it was chosen to deal with inverted values. Secondly, DSP processors are typically efficient at performing add, subtract and multiply operations, but not divisions. Thus the processing should be arranged so as to minimise the number of division operations required.

APPROACH 1

Conversion of the MDE algorithm described in Appendix A – Approach 1 is described below. The modulation depth as per Eq. (A1.1) is inverted so as to never exceed 1.0.

$$MD_n = e_n / P_n \qquad \text{(B1.1)}$$

The Knee point and Limit point are subsequently inverted and Eqs. (A1.2) to (A1.4) can be re-written as follows:

(a) For inverted "signal depths" greater than or equal to the inverted Knee point:

$$MD'_n = MD_n^{X_{MD}} \quad \text{for} \quad MD_n \leq K_{MD} \qquad \text{(B1.2)}$$

(b) For inverted "signal depths" less than the inverted Knee point but greater than the inverted Limit point:

$$MD'_n = MD_n \times A + B \qquad \text{(B1.3)}$$

for $K_{MD} < MD_n < L_{MD}$ and $K_{MD}^{X_{MD}} < L_{MD}$

where $A = (L_{MD} - K_{MD}^{X_{MD}}) / (L_{MD} - K_{MD})$

and $B = L_{MD} \times (1 - A)$

(c) For inverted "signal depths" less than the inverted Limit point the envelope signal level is preserved.

$$MD'_n = MD_n \quad \text{for} \quad MD_n \geq L_{MD} \tag{B1.4}$$

The modified trough level (Eq. A1.5) can be expressed as a function of the modified inverted modulation depth (MD'_n).

$$T'_n = P_n \times MD'_n \tag{B1.5}$$

APPROACH 1.A

For Approach 1.A, the modified envelope signal is defined as:

$$e'_n = e_n \times C_n + D_n \tag{B1.6}$$

where $C_n = (P_n - T'_n) / (P_n - T_n)$ and $D_n = P_n \times (1 - C_n)$

Note however that the constant C_n will always be greater than or equal to 1.0. Inverting C_n will require a divide operation for each calculation of e'_n, thus it would be more efficient to scale C_n (and thus D_n) by a factor of 1/S (where $S = 2^{12}$ for a 24-bit DSP) when storing these constants and inverse scaling of S can be applied to Eq. (B1.6) as shown in Eq. (B1.7). It is efficient to scale by a power of 2 because this can typically be performed using a barrel right or left shift operation in a DSP rather than a divide or multiply operation respectively.

$$e'_n = (e_n \times C'_n + D'_n) \times S \tag{B1.7}$$

where $C'_n = (1/S) \times (P_n - T'_n) / (P_n - T_n)$ and $D'_n = P_n \times (1/S - C'_n)$

APPROACH 1.B

For Approach 1.B, the constants C'_n and D'_n used in Eq. (B1.7) can be defined as follows.

$$C'_n = (1/S) \times (M_n - T'_n) / (M_n - T_n) \quad \text{and} \quad D'_n = M_n \times (1/S - C'_n)$$

APPROACH 2

APPROACH 2.A

Conversion of the MDE algorithm described in Appendix A – Approach 2.A is described below. The "signal depth" as per Eq. (A2.1) is inverted so as to never exceed 1.0.

$$sd_n = e_n / P_n \qquad \text{(B2.1)}$$

The Knee point and Limit point are subsequently inverted and Eqs. (A2.2) to (A2.4) can be re-written as follows:

(a) For inverted "signal depths" greater than or equal to the inverted Knee point:

$$e'_n = P_n \times sd_n^{X_{MD}} \qquad \text{for} \quad sd_n \geq K_{MD} \qquad \text{(B2.2)}$$

(b) For inverted "signal depths" less than the inverted Knee point but greater than the inverted Limit point:

$$e'_n = P_n \times (sd_n \times A + B) \qquad \text{(B2.3)}$$

for $K_{MD} > sd_n > L_{MD}$ and $K_{MD}^{X_{MD}} > L_{MD}$

where $A = (L_{MD} - K_{MD}^{X_{MD}}) / (L_{MD} - K_{MD})$

and $B = L_{MD} \times (1 - A)$

(c) For inverted "signal depths" less than the inverted Limit point the envelope signal level is preserved.

$$e'_n = e_n \qquad \text{for} \quad sd_n \le L_{MD} \qquad \text{(B2.4)}$$

APPROACH 2.B

Equations (B2.2) to (B2.4) can be employed with the restriction that the "signal depth" and thus the modified envelope signal level is only calculated when the envelope signal level is less than the mid-point level.

Alternatively, distortion may be alleviated by re-defining the "signal depth" as per Eq. (A2.6). Again the "signal depth" must be inverted so as to never exceed 1.0.

$$sd_n = e_n \,/\, (2 \times M_n - e_n) \qquad \text{for} \quad e_n < M_n \qquad \text{(B2.5)}$$

The Knee point and Limit point are subsequently inverted and Eqs. (A2.7) to (A2.9) can be re-written as follows:

(a) For inverted "signal depths" greater than or equal to the inverted Knee point:

$$e'_n = M_n \times sd_n^{(X_{MD}-1)} \qquad \text{for} \quad sd_n \ge K_{MD} \;\&\; e_n < M_n \qquad \text{(B2.6)}$$

(b) For inverted "signal depths" less than the inverted Knee point but greater than the inverted Limit point:

$$e'_n = M_n \times (sd_n \times A + B) \qquad \text{(B2.7)}$$

for $K_{MD} > sd_n > L_{MD}$ & $K_{MD}^{X_{MD}} > L_{MD}$ & $e_n < M_n$

where $$A = \frac{2 \times (L_{MD} - K_{MD}^{X_{MD}})}{(L_{MD} - K_{MD}) \times (1 + L_{MD}) \times (1 + K_{MD}^{X_{MD}})}$$

and $$B = L_{MD} \times (2 / (1 + L_{MD}) - A)$$

(c) For inverted "signal depths" less than the inverted Limit point the envelope signal level is preserved.

$$e'_n = e_n \qquad \text{for} \quad sd_n \le L_{MD} \qquad \text{(B2.8)}$$

REFERENCES

Abberton, E., & Fourcin, A. (1978). "Intonation and speaker identification," Lang. Speech **21**, 305-318.

Brokx, J. P. L., & Nooteboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics, **10**, 23-36.

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). "The perception of Contonese lexical tones by early-deafened cochlear implantees," J. Acoust. Soc. Am. **111**, 2250-2256.

Highnam, C., & Morris, V. (1987). "Linguistic stress judgments of language learning disabled students," J. Commun. Disord. **20**, 93-103.

Liberman, P., & Michaels, S. B. (1962). "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech," J. Acoust. Soc. Am. **34**, 922-927.

Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., & Cheung, D. M. C. (2002). "Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children," Int . J. Ped. Otolaryngol. **63**, 137-147.

McDermott, H. J., & Vandali, A. E. (1991). "Spectral Maxima Sound Processor," Australian Patent, 657959; US Patent 788591.

McDermott, H. J., McKay, C. M., & Vandali, A. E. (1992). "A new portable sound processor for the University of Melbourne / Nucleus Limited multielectrode cochlear implant," J. Acoust. Soc. Am. **91**, 3367-3371.

McKay, C. M., McDermott, H. J., & Clark, G. M (1994). "Pitch percepts associated with amplitude-modulated current pulse trains by cochlear implantees," J. Acoust. Soc. Am. **96**, 2664-2673.

McKay, C. M., & McDermott, H. J. (1995). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites," J. Acoust. Soc. Am. **100**, 1081-1092

Moore, B. C. J. (1995). "Hearing" in *The handbook of Perception and Cognition (2nd ed.)*, edited by B. C. J. Moore (Academic Press, Inc., London), pp 267-295.

Nooteboom, S. (1997). "The prosody of speech: Melody and rhythm," in *The handbook of Phonetic Sciences*, edited by W.J. Hardcastle and J. Laver (Blackwell, Oxford), pp 640-673.

Seligman, P. M., Dowell, R. C., Blamey, P. J. (1992). "Multi Peak Speech Procession," US patent 5,095,904.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., & Beiter, A. L. (1991). "Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and the Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant," Ear and Hearing, **12**, 3-22.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. A., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., & Beiter, A. L. (1994). "Evaluation of a new spectral peak (SPEAK) coding strategy for the Nucleus 22 channel cochlear implant system," The Am. J. Otology, **15**, (Suppl. 2), 15-27.

Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (2000). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 Cochlear implant system," Ear & Hearing, **21**, 608-624.

Wells, B., Peppe, S., & Vance, M. (1995). "Linguistic assessment of prosody," in *Linguistics in Clinical Practice*, edited by K. Grundy (Whurr, London), pp 234-265.

Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., Clark, G. M. (1995). "Evaluation of the Nucleus Spectra 22 processor and new speech processing strategy (SPEAK) in postlinguistically deafened adults," Acta Oto-laryngologica (Stockholm), **115**, 629-637.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," Nature, **352**, 236- 238.

Wherein the following Appendix C figures referenced are numbered 1, 2 and 3, it should be noted that any reference to figure 1 should be taken as a reference to figure 10, any reference to figure 2 should be taken as a reference to figure 11 and any reference to figure 3 should be taken as a reference to figure 12.

APPENDIX C: PHASE ALIGNMENT FOR VOCODER BASED SPEECH SYSTEMS

FIELD OF THE INVENTION

This invention relates to the processing of signals derived from sound stimuli, particularly for the generation of stimuli in auditory prostheses, such as cochlear implants and hearing aids, and in other systems requiring vocoder based sound processing or encoding.

BACKGROUND OF THE INVENTION

Voice pitch information can play an important role in speech perception as it provides cues to linguistic features such as intonation (question – statement contrast) and word emphasis (Highnam, & Morris 1987; Nooteboom, 1997; Wells, Peppe, & Vance, 1995) and also to paralinguistic features such as speaker identification and the emotional state of the speaker (Abberton & Fourcin, 1978; Liberman, & Michaels, 1962) and segregation of concurrent speakers (Brokx, & Nooteboom, 1982). Most importantly, voice pitch information is crucial for perception of tonal languages, such as Mandarin and Cantonese, where a change in fundamental voicing frequency within the same phonemic segment causes a change in lexical meaning (Lee *et. al.*, 2002; Ciocca *et. al.*, 2002). Pitch information is also of important to the appreciation of music where the frequency of the fundamental and its harmonics governs the pitch of the signal (Moore 1995).

Various speech processing strategies have been developed for processing of sound signals for use in stimulating auditory prostheses, such as cochlear prostheses and hearing aids. The multi-peak strategy (Seligman, Dowell, & Blamey, 1992; Skinner *et. al.*, 1991) focused particularly on coding of aspects of speech, such as formants and the fundamental voicing frequency. For this strategy voice pitch information was predominantly coded by way of the electrical stimulation rate. Other strategies relied more on general channelization of the

sound signal, such as the Spectral Maxima Sound Processor (SMSP) strategy, which is described in greater detail in Australian Patent No. 657959 and US Patent No. 5597390 by Hugh McDermott and the present applicant. For this strategy voice pitch information (for a voicing frequency below approximately 200 Hz) is generally coded by way of amplitude modulation, at a frequency equal to or related to the voicing frequency, in the envelope signals of each channel.

It is an object of the present invention to provide a sound processing strategy to assist in perception of voice pitch and musical tone in the sound signal.

SUMMARY OF THE INVENTION

Broadly, the present invention provides a method of processing channelised sound signals, wherein the timing of signals in at least some channels is adjusted so that the phase differences between modulated signals across these channels are reduced.

According to one aspect, the present invention provides a sound processing process including at least the steps of:

(a) receiving a sound signal;

(b) processing said sound signal so as to produce a set of channel signals in spaced frequency channels; and

(c) determining an envelope for each channel signal, said envelope being determined so as to retain information about the fundamental frequency;

(d) determining temporal peaks in the envelope of at least selected channel signals, said peaks being related to the fundamental frequency;

(e) selectively applying a timing offset to adjust the timing of the peaks in said selected channel signals, said adjustment being made in response to a predetermined instruction set, so that the phase differences between the peak values in at least said selected channel signals are reduced; and

(f) performing further processing upon at least some of the set of channel signals.

According to another aspect, the present invention provides a sound processing process, including at least the steps of:

(a) receiving a sound signal;

2004242561 24 Mar 2005

(b) processing said sound signal so as to produce a set of channel signals in spaced frequency channels;

(c) Determining a channel envelope for each channel signal, said envelope being determined so as to retain information about the fundamental frequency of said sound signal;

(d) determining a reference signal corresponding to the sum of channel envelopes in a plurality of channels;

(e) Determining the timing of temporal peaks in each of said channel envelopes and in said reference signal, said peaks being related to the fundamental frequency of said sound signal;

(f) Determining a timing offset for at least some of said channels, using a predetermined instruction set, by reference to at least the difference in timing between the peaks in said reference signal and the corresponding peaks in each channel envelope,

(g) Adjusting the timing of the channel signal in said selected channel signals in accordance with the timing offsets, so that the phase differences between the peak values in at least said selected channel signals are reduced;

(h) Determining subsequent reference signals from a sum of the time-shifted signals in said channels, and

(i) Using the time adjusted channels signals for further processing.

Experiments conducted with users of cochlear implant prostheses have indicated that the frequency of amplitude modulated electrical signals can be reliably detected when the modulation depth is sufficiently deep (McKay, McDermott, & Clark, 1994; and work in progress by the present inventors). Results between users were varied but showed consistent identification for depths of at least 6 to 12 dB (however a few users have been found to require a deeper modulation depth of approximately 30 dB which is typically equivalent to their full electrical dynamic range). Channelization of the sound signal, as is done by most vocoder based speech coding strategies today, provides an estimate of the narrow-band envelope signal in each band-pass filter channel. For voiced speech signals, voicing pitch (periodicity) information will be represented in the narrow-band envelope signals as an amplitude-modulated signal with a frequency equal to or related to the voicing frequency.

These amplitude-modulated narrow-band envelope signals may however differ in phase between channels. Experimental findings with users of cochlear implant prostheses have shown that phase differences between modulation signals in nearby channels (up to 3 to 5 electrode channels apart or approximately 2 to 4 mm apart) can compromise identification of modulation frequency due to either or both i) temporal integration of signals across channels (McKay & McDermott, 1996; and work in progress by the present applicant), or ii) spread of the electrical stimulus current field. It was thus proposed that minimisation of the phase differences between peaks in amplitude-modulated envelope signals across channels might provide better perception of the modulation frequency.

As with most multi-channel speech processing systems, the preferred implementation of the present invention provides a sound processing device having means for channelization of the input sound into a plurality of spaced frequency channels by using a bank of band-pass filters; means for estimating the narrow-band envelope signal (which include both the slow envelope components of less than 50 Hz and fine structure of greater that 50 Hz up to approx 400 Hz) of the signal in each spaced frequency channel. In particular, the system provides means for minimising the phase offset (time difference between temporal peaks) in the envelope signals of each channel.

The present invention encompasses reducing the phase differences between channel signals across all channels using a common reference signal, reducing the phase differences between some channels and not others (for example aligning only channels within a certain frequency range), and aligning channels within different frequency bands to different reference signals.

Furthermore, the invention encompasses both the reference signal based approach which is described in detail, and alternative approaches to reducing phase differences between channels.

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the invention may be more readily understood, one presently preferred embodiment of the invention will now be described with reference to the accompanying drawings in which:

2004242561 24 Mar 2005

Figures 1 is a schematic representation of the overall signal processing arrangement;

Figure 2 is a schematic representation of the phase alignment algorithm;

Figure 3 is a schematic representation of the time shift processing system; and

Figure 4 displays comparative electrodograms (stimulus output patterns) of amplitude modulated sound signals to show the overall effect of the implementation.

DESCRIPTION

According to the preferred implementation, in order to minimise the phase difference between channels it is important to establish a reference envelope signal to which the phase of the envelope signals in each channels can be aligned. Ideally, the modulation frequency of the reference signal should correspond to the fundamental voicing frequency of the signal and should be robust to the effects of competing noise. Next, a phase offset (or time shift) for each channel could be determined so as to best align the peaks in the temporal envelope of each channel with that of the reference signal. The time shift can be up to half the lowest expected voicing/modulation period (for 80 Hz this is approximately 6 ms) and can be a positive or negative time shift (i.e. ± 6 ms). A temporal peak detection algorithm is preferably used to determine the location in time of maximum turning points (peaks) in the envelope signal. Small peaks or peaks too close in time should be ignored. Having established the peak times for a channel, the peak times could be aligned (if possible) with those of the reference signal by introducing a time shift to the channel in question.

Alternatively a more brute force approach might be employed that determines the optimum time shift based on an analysis of the combination (product or summation or similar rule) of the reference signal and the channel in question for all possible time shifts. Finally, the phase offset (time shift) should be applied to each channel in such a way so as to minimise any distortion of the envelope signals resulting from the time shift.

The reference signal in a preferred form is generated using an iterative procedure. Initially, all filter bank envelope signals are summed together over some finite time window. Given a positive signal-to-noise ratio we would expect,

at minimum, that some of the channels contain periodicity information pertaining to the fundamental voicing/modulation frequency. This reference signal is used to adjust the phase offset (time shift), using the techniques discussed above, of each channel in turn for the given time window. Processing would then begin again for the next time window/pass except in this case the reference signal would be constructed by summing the time shifted channel signals together. Distortion in each channel due to dynamic adjustment of the time shift could be minimised by low-pass filtered the time shift and/or only applying it at time points for a given channel where its envelope signal is at a local minimum (temporal trough).

Provided that the fundamental voicing (or modulation) frequency of the entire signal is steady we would expect that within a few iterations (time steps), the phase offset (time shift) for each channel should converge on some value that aligns the temporal peaks of each channel. Furthermore, provided that the fundamental voicing or modulation frequency does not change too rapidly the algorithm should be able to adapt and follow the changes in frequency over time. For unvoiced (non-periodic) signals we would expect that the phase offset for each channel would vary randomly. However given that the location in time of temporal peaks for unvoiced signals varies randomly anyway we would not expect this to be a problem. For voiced (periodic) signals combined with noise, we would expect that this algorithm will align any periodic temporal patterns available in the envelope signals of each channel but that errors in alignment will be inversely proportional to some function of the signal-to-noise ratio.

The present invention may be implemented using alternative methods for generating or applying the reference signal, provided that the broad functional intention of supplying additional information about the fundamental frequency (F0) and reducing phase differences between channels is met.

Referring to Figures 1 to 3, the presently preferred embodiment of the invention is described with reference to its use with the Advanced Combinational Encoder (ACE) strategy (Vandali *et. al.*, 2000). Note, however that it could equally be applied to other speech coding strategies such as the Continuous Interleaved Sampling (CIS) strategy (Wilson *et. al.*, 1991), or any other system in which input sound signals are channelised before further processing. Although

the present invention is principally described with reference to a cochlear implant system, the invention is not limited to such an implementation.

As with the ACE strategy, electrical signals corresponding to sound signals received via a microphone 1 and pre-amplifier 2 are processed by a bank of N parallel filters 3 tuned to adjacent frequencies (typically N = 20 for the ACE strategy). Each filter channel includes a band-pass filter 31 and an envelope detector 32 to provide an estimate of the narrow-band envelope signal in each channel. The band-pass filters are typically narrow (approximately 180 Hz wide – 3 dB bandwidth) for apical (low-frequency) channels and increase in bandwidth (typically up to 1000 Hz or more) for more basal (higher frequency) channels. The envelope detectors 32, which effectively comprise a full-wave (quadrature) rectifier followed by a low-pass filter, typically pass fundamental (modulation) frequency information up to approximately 200 Hz but for some implementations (including this preferred embodiment) can accommodate frequencies as high as approximately 400 Hz.

The filter bank is used to provide an estimate of the envelope signals in each channel at regular time intervals known as the analysis or update rate. The ACE strategy can employ an update rate which can be adjusted from as low as approximately 200 Hz up to as high as approximately 4000 Hz (depending on the hardware device used). In the present invention, an update rate of approximately 1200 Hz (or 1600 Hz) is employed so that modulation frequencies of approximately 300 Hz (or 400 Hz) can be adequately sampled (note, amplitude modulation identification experiments with users of cochlear implant prostheses have indicated that update/stimulation rates of at least four times the modulation frequency are required for adequate analysis/coding of the signal, McKay, McDermott, & Clark, 1994).

It is desirable that for each intended application, the envelope is determined so as to retain information about the fine temporal structure in the range of fundamental frequencies for the intended application. In the context of speech, the range covering the voicing frequency(F0) should be captured. Otherwise, the desired pitch information will not be supplied to the user.

The outputs of the N-channel filter bank are then processed by the phase alignment algorithm which is used to minimises the phase difference (or align

temporal peaks) between amplitude-modulated envelope signals across channels, prior to further processing by the speech coding strategy. This can be carried out for all channels in the system, or at least for channels in the voiced frequency range (i.e. first, second and third formant frequency range which ranges from approximately 100 Hz up to 3 to 4 KHz).

The phase alignment algorithm is depicted by blocks 4, 5 and 6 of Figure 1 and in greater detail in Figures 2 and 3. The channel envelope signals derived from the filter bank 3 are stored in time buffers 4 which are used to construct a reference signal 5 and to calculate appropriate time shifts 6 for each channel so as to align temporal peaks in the reference signal to those of channel signals.

The time shifts calculated are applied to the envelope signals of each channel. These time shifted envelope signals effectively replace the original envelope signals derived from the filter bank and processing then continues as per the original speech coding strategy. For the ACE strategy M of the N channels having the largest amplitude at a given instance in time are selected 7 (typically M = 8 for this embodiment). The M selected channels are then used to generate M electrical stimuli 8 corresponding in stimulus intensity and electrode number to the amplitude and frequency of the M selected channels. These M stimuli are transmitted to the Cochlear implant 10 via a radio-frequency link 9 and are used to activate M corresponding electrode sites. This algorithm may be applied to the channelised sound signal, and subsequent processing continue as per any chosen processing strategy for the cochlear implant. This strategy is specific to this stage of processing, and hence is applicable to any strategy, which employs channelisation and subsequent processing (with modifications as may be dictated by the requirements of the selected strategy).

It will be understood that alternative techniques could be used to determine the time shifts, and to apply them to the channel signals, within the scope of the present invention.

The first task of the phase alignment algorithm is to construct a reference signal 5 for which envelope channel signals can be aligned to. The duration of the reference signal should be sufficiently long enough to contain at least two cycles of the modulation signal (i.e at least two temporal peaks). Setting a lower limit of approximately 80 Hz for fundamental voicing frequencies to be analysed by this

2004242561 24 Mar 2005

system dictates a minimum duration of 2 ×1/80 = 25 milliseconds (ms) for the reference signal buffer 52 length. In this implementation, initially the time shifts (phase offsets) for all channels are initialised to zero and the reference signal is generated by summing the channel envelope signals together 51 for each time point in the reference signal buffer. However, in subsequent time frames the time shifts will vary and thus will result in construction of the reference signal from time shifted channel envelope signals. The time shift can be up to half the lowest expected voicing/modulation period (for 80 Hz this is approximately 6 ms) and can be a positive or negative time shift (i.e. ± 6 ms). Thus, the channel envelope signal buffers 41 which hold the channel envelope signals used to construct the reference signal must be at least the length of the reference signal buffer (25 ms) plus twice the maximum time shift (2 × 6 ms) which equals 37 ms.

When constructing the reference signal from the channel envelope signals, any channels that contain modulated signals that are aligned in phase with respect to one another will sum constructively so as to enforce the modulated signal in the reference signal. In contrast, channels that contain modulated signals that are out of phase with one another will tend to diminish or introduce additional peaks in the reference signal. Modulated signals that are 180 degrees out of phase will be the most destructive. For voiced speech input signals we would expect that the reference signal (i.e. combination of all channels) will contain modulation terms that are related to the F0 of the signal but that the strength of these terms will vary with characteristics of the input signal and the band-pass filter bank used to analyse the input signal. We might also expect that higher frequency modulation terms may be present in the reference signal. The relative amplitude of the F0 signal to the higher frequency modulation terms will also vary with characteristics of the input signal and the band-pass filter-bank. However, in subsequent time frames, if the time shifts for each channel are adjusted so as to align temporal peaks then this will result in strengthening of the F0 terms in the reference signal.

The next task of the phase alignment algorithm is to determine appropriate time shifts 6 for each channel. Block 6 in Figure 3 depicts the processing path for this operation for one channel envelope signal (n). A peak detector 61 is used to determine the relative time-location of all temporal peaks in the reference signal

2004242561 24 Mar 2005

52 by finding all maximum turning points in the signal. Setting an upper limit of approximately 330 Hz for the fundamental voicing frequencies to be analysed by this system suggests that temporal peaks closer than 1/330 = 3 ms should be ignored, or rather the smaller of the two peaks should be discarded. If two or more temporal peaks in the reference signal (over a 25 ms period) are located (i.e. a modulation frequency equal to or higher than 80 Hz is present) then calculation of the phase offset (or time shift) for the channel envelope signals that best aligns their peaks in time to those of the reference signal are determined.

This is carried out for each channel by determining the time-location of all temporal peaks (excluding those closer than 3 ms apart) in the channel envelope signal 62 and determining the time shift required for each temporal peak that will align it with the peaks located in the reference signal. Specifically, for each of the peaks located by the channel peak detector 62, the channel envelope signal is shifted in time 63 so as to align the peak with a peak in the reference signal. A correlation, of sorts, between the reference signal and the shifted channel envelope signal is then performed 64, 65, 66. The correlation is carried out by multiplying 64 each time point in the time-shifted channel envelope signal 63 by each time point in the reference signal 51 and summing 66 the multiplied terms 65. These operations are carried out consecutively for all peaks in the channel envelope signal and in the reference signal. An optimal time shift is then determined 67 by selecting the smallest time shift (from 62), compared to the current time shift, which provides the largest correlation term (derived from 66). These operations are carried out for all channels and the optimal time shifts for are then used to shift the channel envelope signal buffers 41 in time. The time shifted channel envelope signals from block 4 are referenced from a time point mid-way through the channel envelope buffers 41 plus or minus the time shift. Thus an average delay of half the channel envelop buffer length (37/2 = 18.5 ms) will be introduced by this algorithm.

To illustrate the effect of the phase alignment algorithm on the coding of amplitude modulated sound signals, stimulus output patterns, known as electrodograms (which are similar to spectrograms for acoustic signals), which plot stimulus intensity (plotted as log current level) for each electrode (channel) as a function of time, were recorded for the ACE strategy shown in Fig. 4(a) and for

2004242561 24 Mar 2005

application of the phase alignment to the ACE strategy shown in Fig. 4(b). The sound signal used in these recordings consisted of two sinusoidal amplitude modulated (SAM) pure tones of frequencies 600 and 1200 Hz respectively. The modulation frequencies for both SAM tones was 100 Hz and the modulation phase difference between SAM tones was 180 degrees. For the standard ACE strategy in Fig. 4(a), it can be seen that the envelope of the stimulus signals on electrodes 13 to 16 are 180 degree out-of-phase with those of electrodes 18 to 21. In contrast with application of the phase alignment algorithm to the ACE strategy in Fig. 4(b), it can be seen that all electrodes carrying periodic modulation are now in phase. Note, the stimulus signal on electrode 17 is not very periodic and is not and is not affected much by the algorithm.

Some additional processing/optimisations can also be carried out to improve the algorithm's performance. For instance:

(i) distortion in channel envelope signals can be reduced by only applying an adjusted time shift value when the channel envelope signal level is at a trough (valley) or by low-pass filtering the time shift values.

(ii) Interactions between harmonic frequencies and the frequency boundaries imposed by the filter bank result in envelope channels signals that contain periodic signals that follow F0 but contain additional temporal peaks. These additional peaks lie between those of the underlying F0 pattern and introduce higher frequency modulation terms (e.g. 2 or 3 times F0). In addition, the presence of noise will also introduce temporal peaks in the envelope signals, which are unrelated to the F0 periodicity.

One approach to reducing some of these effects is to smooth (low-pass filter, LPF) the reference signal using an integration window (finite impulse response filter) of length related to an estimate of the fundamental voicing frequency (F0). Alternatively, a fixed LPF could be employed with a cut-off frequency of approximately 100 Hz that will attenuate the effect of higher order harmonics when F0 is low (i.e. around 80-150 Hz).

(iii) Many of the processing blocks can be bypassed when the signal level is small (e.g. below some threshold in which the effect of the algorithm will go unnoticed);

(iv). Processing time can be reduced by calculating the time shifts (blocks 5 and 6) at a lower frequency (e.g. decimation of the time shift operations by a factor of 2 or 4).

It will be appreciated that the present invention can be applied to all or only some of the channel signals. For example, a reference signal may be generated based only upon a band of channels, and used only to modify the phase of those signals. An example of this may be to modify phase only for signals in a range corresponding to voiced signals, and not modify other channels. In this instance, appropriate buffer times will be required to compensate for processing delays in other channels.

Since modifications within the spirit and scope of the invention may be readily effected by persons skilled in the art, it is to be understood that the invention is not limited to the particular embodiment described above.

REFERENCES

Abberton, E., & Fourcin, A. (1978). "Intonation and speaker identification," Lang. Speech **21**, 305-318.

Brokx, J. P. L., & Nooteboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics, **10**, 23-36.

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). "The perception of Contonese lexical tones by early-deafened cochlear implantees," J. Acoust. Soc. Am. **111**, 2250-2256.

Highnam, C., & Morris, V. (1987). "Linguistic stress judgments of language learning disabled students," J. Commun. Disord. **20**, 93-103.

Liberman, P., & Michaels, S. B. (1962). "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech," J. Acoust. Soc. Am. **34**, 922-927.

Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., & Cheung, D. M. C. (2002). "Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children," Int . J. Ped. Otolaryngol. **63**, 137-147.

McDermott, H. J., & Vandali, A. E. (1991). "Spectral Maxima Sound Processor," Australian Patent, 657959; US Patent 788591.

McDermott, H. J., McKay, C. M., & Vandali, A. E. (1992). "A new portable sound processor for the University of Melbourne / Nucleus Limited multielectrode cochlear implant," J. Acoust. Soc. Am. **91**, 3367-3371.

McKay, C. M., McDermott, H. J., & Clark, G. M (1994). "Pitch percepts associated with amplitude-modulated current pulse trains by cochlear implantees," J. Acoust. Soc. Am. **96**, 2664-2673.

McKay, C. M., & McDermott, H. J. (1996). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites," J. Acoust. Soc. Am. **100**, 1081-1092

Moore, B. C. J. (1995). "Hearing" in *The handbook of Perception and Cognition (2nd ed.)*, edited by B. C. J. Moore (Academic Press, Inc., London), pp 267-295.

Nooteboom, S. (1997). "The prosody of speech: Melody and rhythm," in *The handbook of Phonetic Sciences*, edited by W.J. Hardcastle and J. Laver (Blackwell, Oxford), pp 640-673.

Seligman, P. M., Dowell, R. C., Blamey, P. J. (1992). "Multi Peak Speech Procession," US patent 5,095,904.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., & Beiter, A. L. (1991). "Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and the Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant," Ear and Hearing, **12**, 3-22.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. A., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., & Beiter, A. L. (1994). "Evaluation of a new spectral peak (SPEAK) coding strategy for the Nucleus 22 channel cochlear implant system," The Am. J. Otology, **15**, (Suppl. 2), 15-27.

Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (2000). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 Cochlear implant system," Ear & Hearing, **21**, 608-624.

Wells, B., Peppe, S., & Vance, M. (1995). "Linguistic assessment of prosody," in *Linguistics in Clinical Practice*, edited by K. Grundy (Whurr, London), pp 234-265.

Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., Clark, G. M. (1995). "Evaluation of the Nucleus Spectra 22 processor and new speech processing strategy (SPEAK) in postlinguistically deafened adults," Acta Oto-laryngologica (Stockholm), **115**, 629-637.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," Nature, **352**, 236- 238.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A sound processing process including at least the steps of:

(a) receiving a sound signal;

(b) processing said sound signal so as to produce a set of signals in spaced frequency channels; and

(c) performing further processing upon at least some of the set of signals;

wherein said process further includes the step of selectively increasing the modulation depth of the envelope signal for at least selected channels in response to a predetermined instruction set, prior to step (c).

2. A sound processing process according to claim 1, wherein the modulation depth is estimated for a channel after step (b) by calculating the ratio of the peak amplitude to the trough amplitude of the envelope signal over a time duration sufficient to allow for fundamental voicing frequencies to be determined.

3. A sound processing process according to claim 1, wherein the modulation depth is adjusted by reducing the trough amplitude while substantially preserving the value of the peak amplitudes.

4. A sound processing process according to claim 2, wherein the level of increase of modulation depth for a channel is dependant upon the value of the modulation depth.

5. A sound processing process according to claim 4, wherein

- if the modulation depth is below a predetermined value K, the modulation depth is increased by a power function;
- if the modulation depth is greater that K and smaller than a limit point L, then the modulation depth is increased by a linear function; and
- if the modulation depth is greater than L, the modulation depth is not modified.

6. A sound processing process according to claim 1, wherein the modulation depth is adjusted by increasing the value of the peak amplitudes and reducing the value of the trough amplitudes.

7. A sound processing process according to claim 1, wherein the shape of the modulation peaks within the channel are substantially preserved after the change to the modulation depth.

8. A sound processing process according to claim 1, wherein the process is carried out for an auditory prosthesis.

9. A sound processing process according to claim 8, wherein the process is carried out prior to applying a speech processing strategy.

10. A sound processing process according to claim 9, wherein the auditory prosthesis is a cochlear implant.

11. A device operatively adapted to carry out the process of any one of the preceding claims.

10. An auditory prosthesis or part thereof, including software intended to operatively carry out the process of claim 8.

11. A software product, stored on a storage medium, including instructions intended to operatively carry out the process of any one of claims 1-10.

DATED this 23rd day of March 2005

HEARWORKS PTY LIMITED

WATERMARK PATENT & TRADE MARK ATTORNEYS
290 BURWOOD ROAD
HAWTHORN VICTORIA 3122
AUSTRALIA

P23458AU00

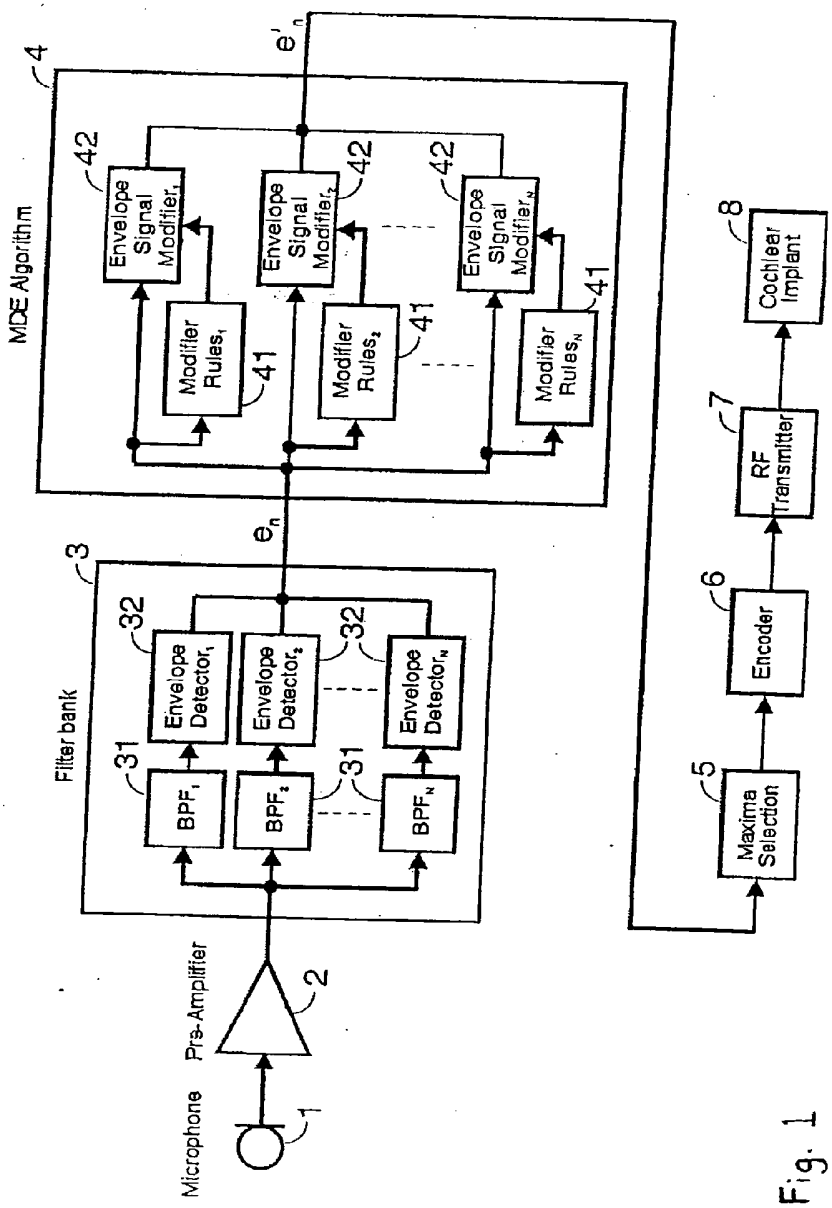
Microphone
1
Pre-Amplifier
2
Filter bank
3
BPF1
BPF2
BPFN
31
Envelope Detector1
Envelope Detector2
Envelope DetectorN
32
en
MDE Algorithm
4
Modifier Rules1
Modifier Rules2
Modifier RulesN
41
Envelope Signal Modifier1
Envelope Signal Modifier2
Envelope Signal ModifierN
42
e'n
Maxima Selection
5
Encoder
6
RF Transmitter
7
Cochlear Implant
8

Fig. 1

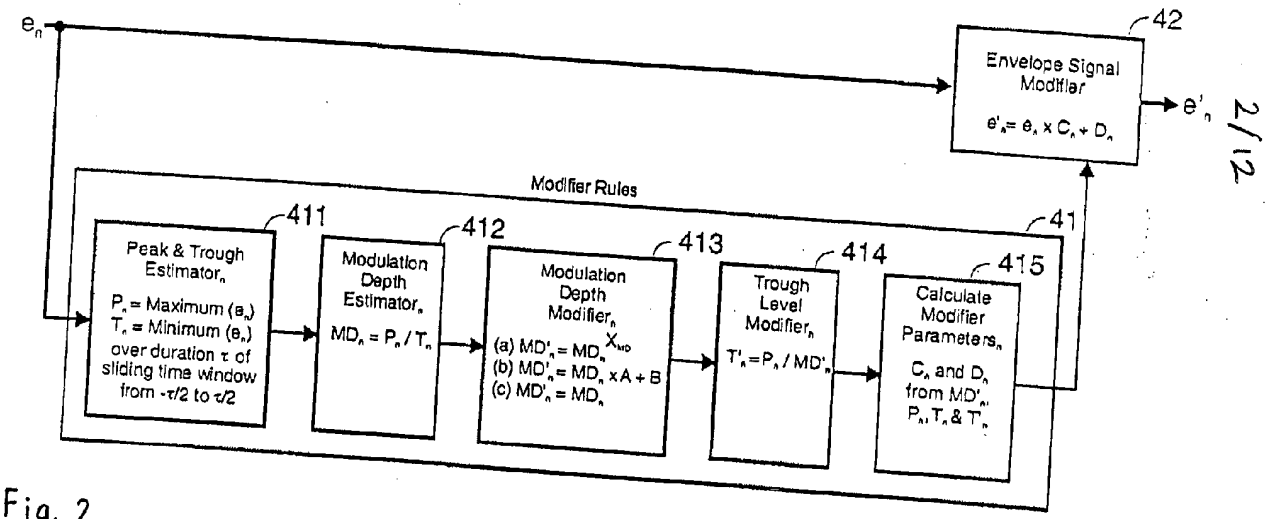
e_n
42
Envelope Signal Modifier
$e'_n = e_n \times C_n + D_n$
e'_n
Modifier Rules
41
411
Peak & Trough Estimator$_n$
P_n = Maximum (e_n)
T_n = Minimum (e_n)
over duration τ of sliding time window from $-\tau/2$ to $\tau/2$
412
Modulation Depth Estimator$_n$
$MD_n = P_n / T_n$
413
Modulation Depth Modifier$_n$
(a) $MD'_n = MD_n^{X_{MD}}$
(b) $MD'_n = MD_n \times A + B$
(c) $MD'_n = MD_n$
414
Trough Level Modifier$_n$
$T'_n = P_n / MD'_n$
415
Calculate Modifier Parameters$_n$
C_n and D_n from MD'_n, P_n, T_n & T'_n

Fig. 2

Input/Output Function for Modulation Depth
where $K_{MD} = 2$, $L_{MD} = 10$ and $X_{MD} = 3$

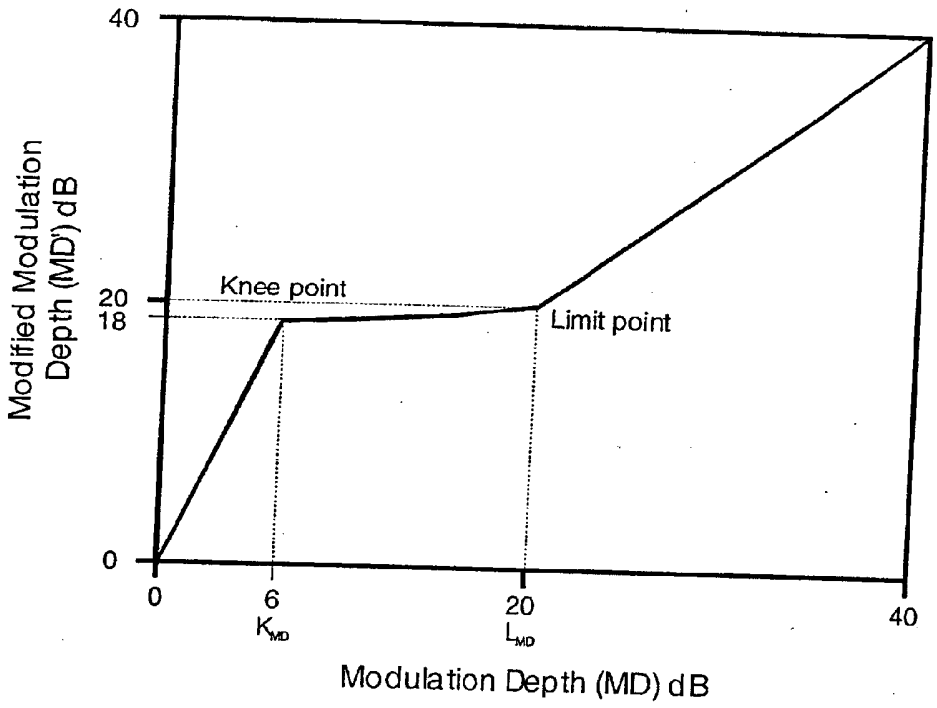

Fig. 3

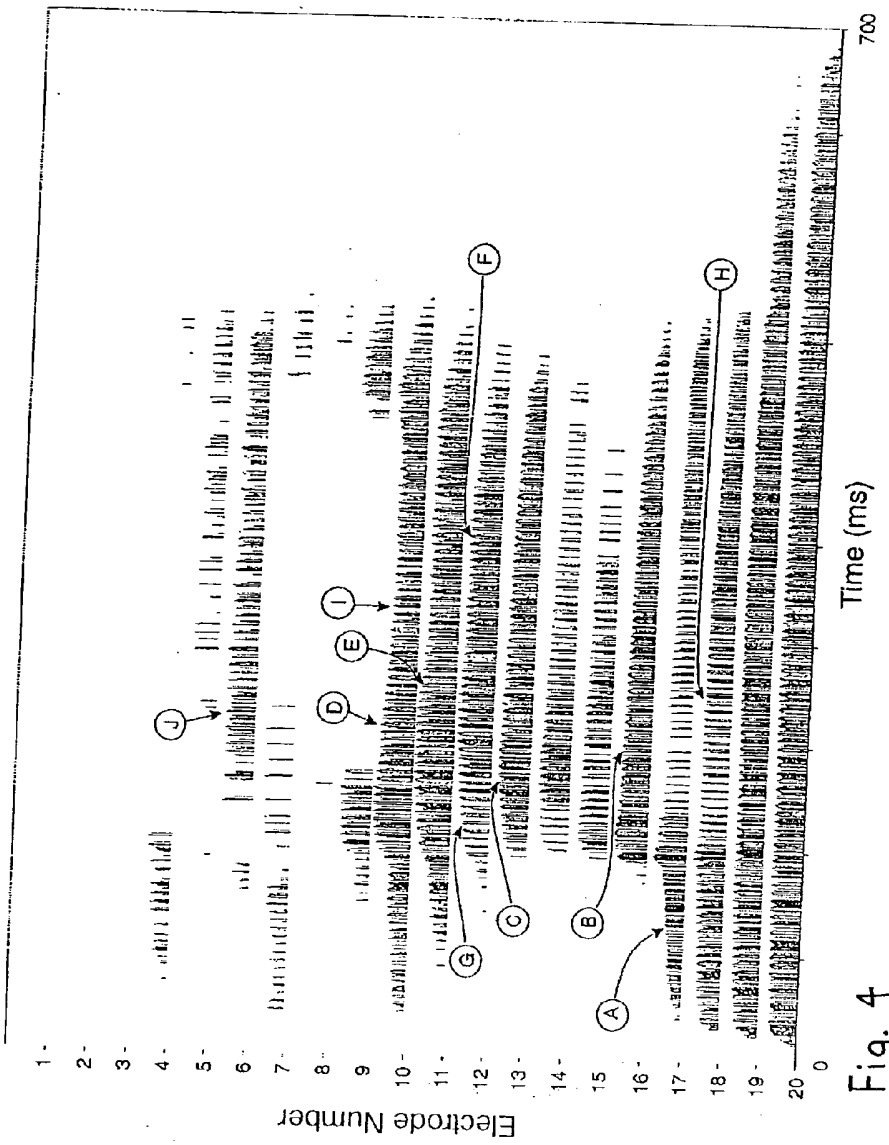

Fig. 4

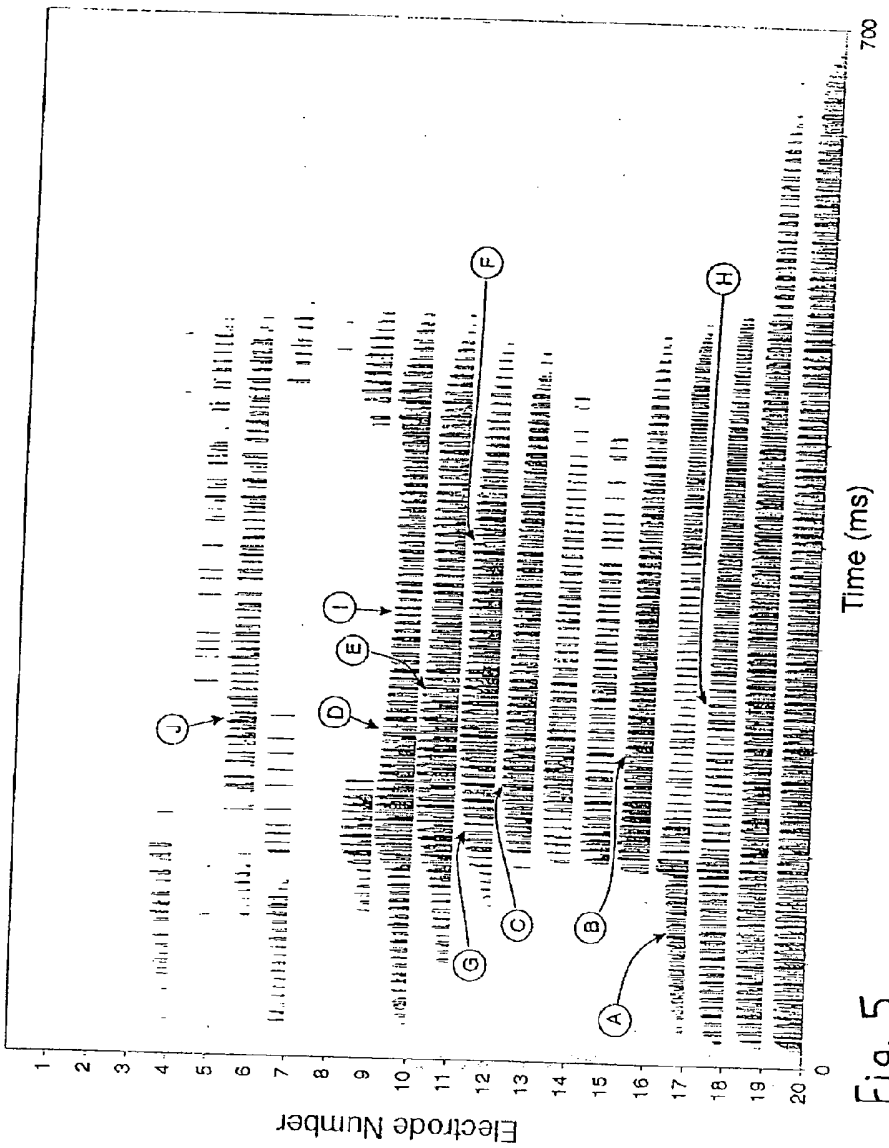

Fig. 5

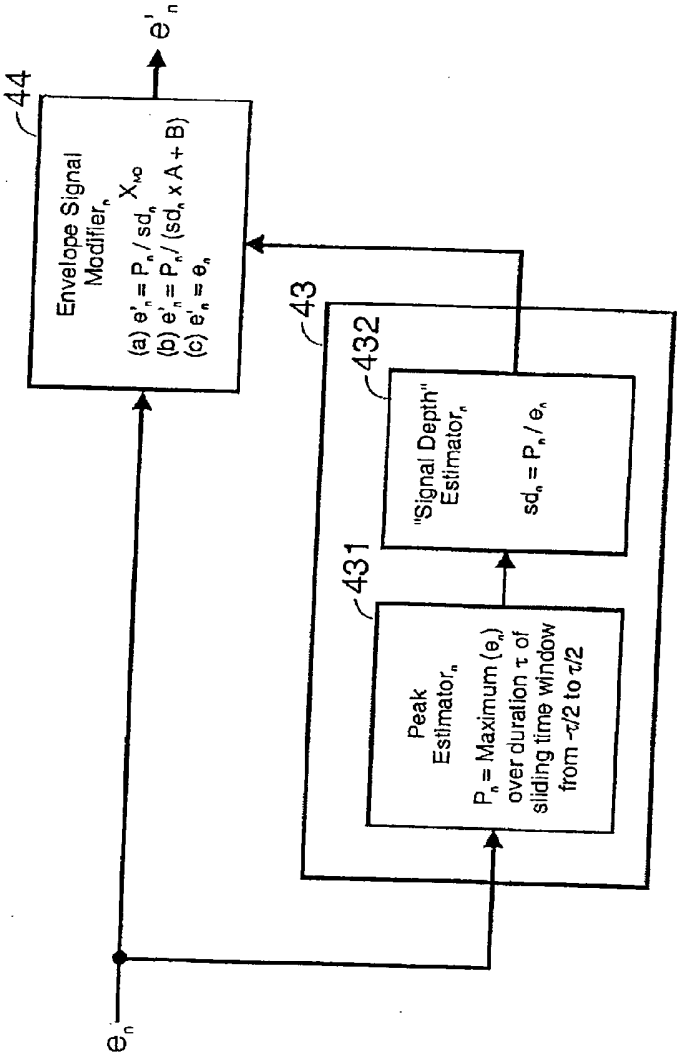
44
Envelope Signal Modifier$_n$
(a) $e'_n = P_n / sd_n$ X_{∞}
(b) $e'_n = P_n / (sd_n \times A + B)$
(c) $e'_n = e_n$
e'_n
43
432
"Signal Depth" Estimator$_n$
$sd_n = P_n / e_n$
431
Peak Estimator$_n$
P_n = Maximum (e_n) over duration τ of sliding time window from $-\tau/2$ to $\tau/2$
e_n

Fig. 6

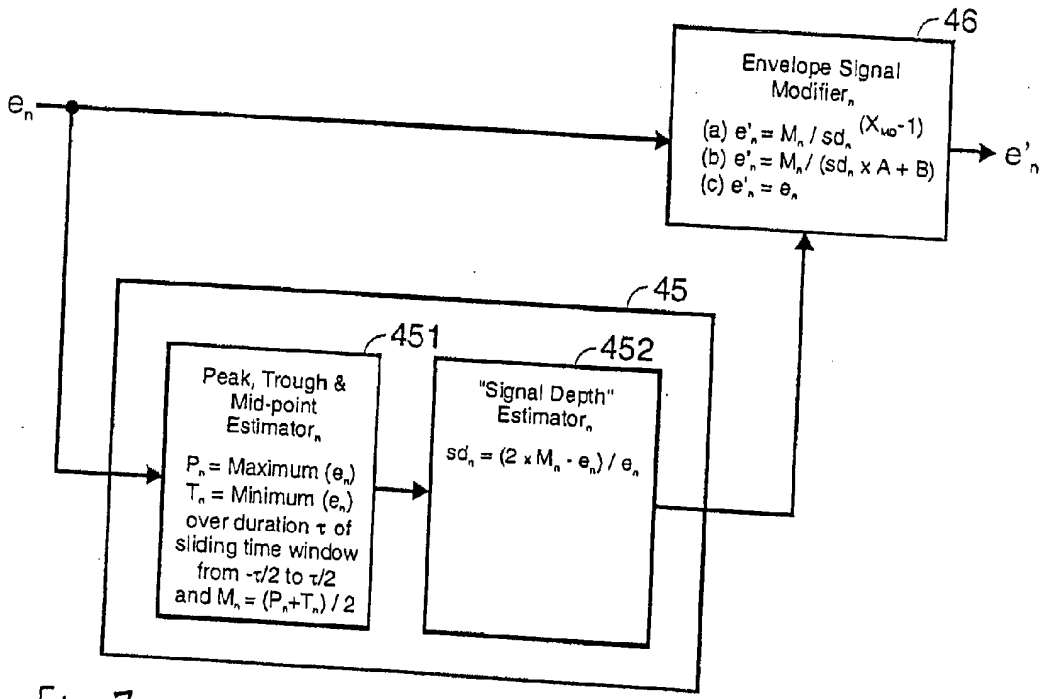
46
Envelope Signal Modifier$_n$
(a) $e'_n = M_n / sd_n^{(X_{MO}-1)}$
(b) $e'_n = M_n / (sd_n \times A + B)$
(c) $e'_n = e_n$
e_n
e'_n
45
451
Peak, Trough & Mid-point Estimator$_n$
P_n = Maximum (e_n)
T_n = Minimum (e_n)
over duration τ of sliding time window from $-\tau/2$ to $\tau/2$
and $M_n = (P_n+T_n) / 2$
452
"Signal Depth" Estimator$_n$
$sd_n = (2 \times M_n - e_n) / e_n$

Fig. 7

7/12

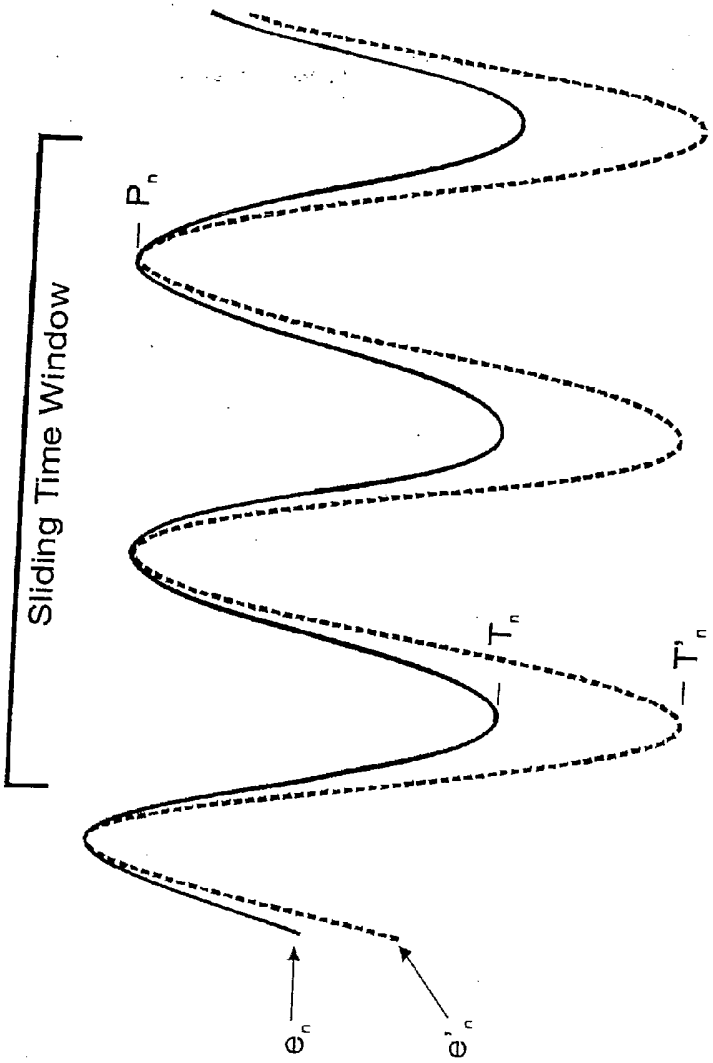

Fig. 8

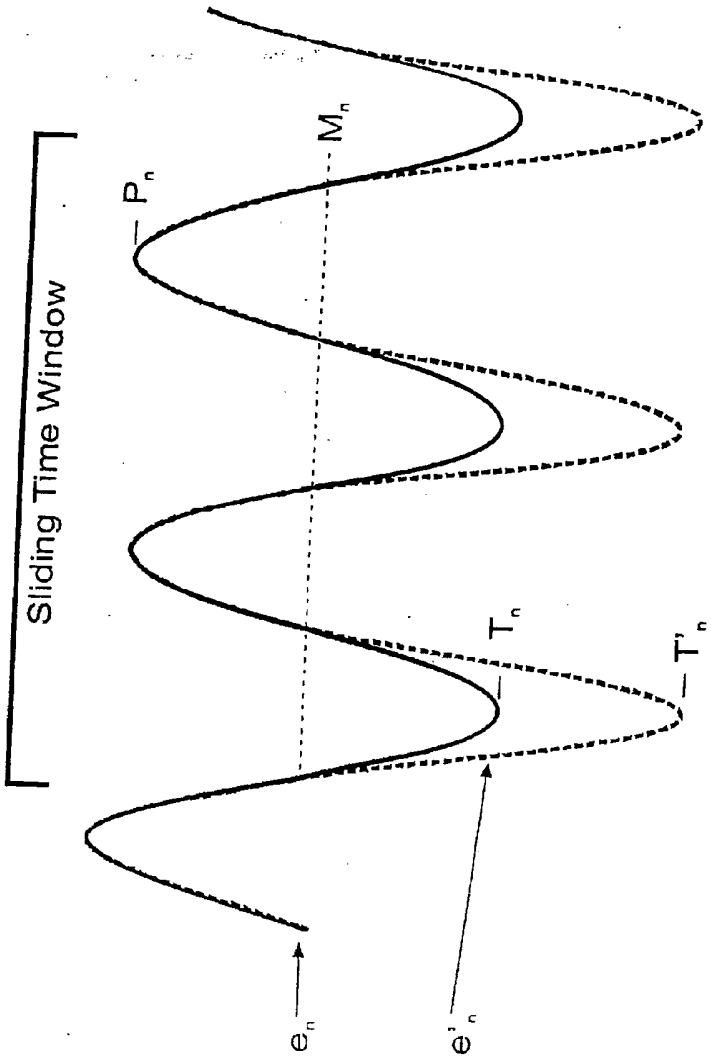

Fig. 9

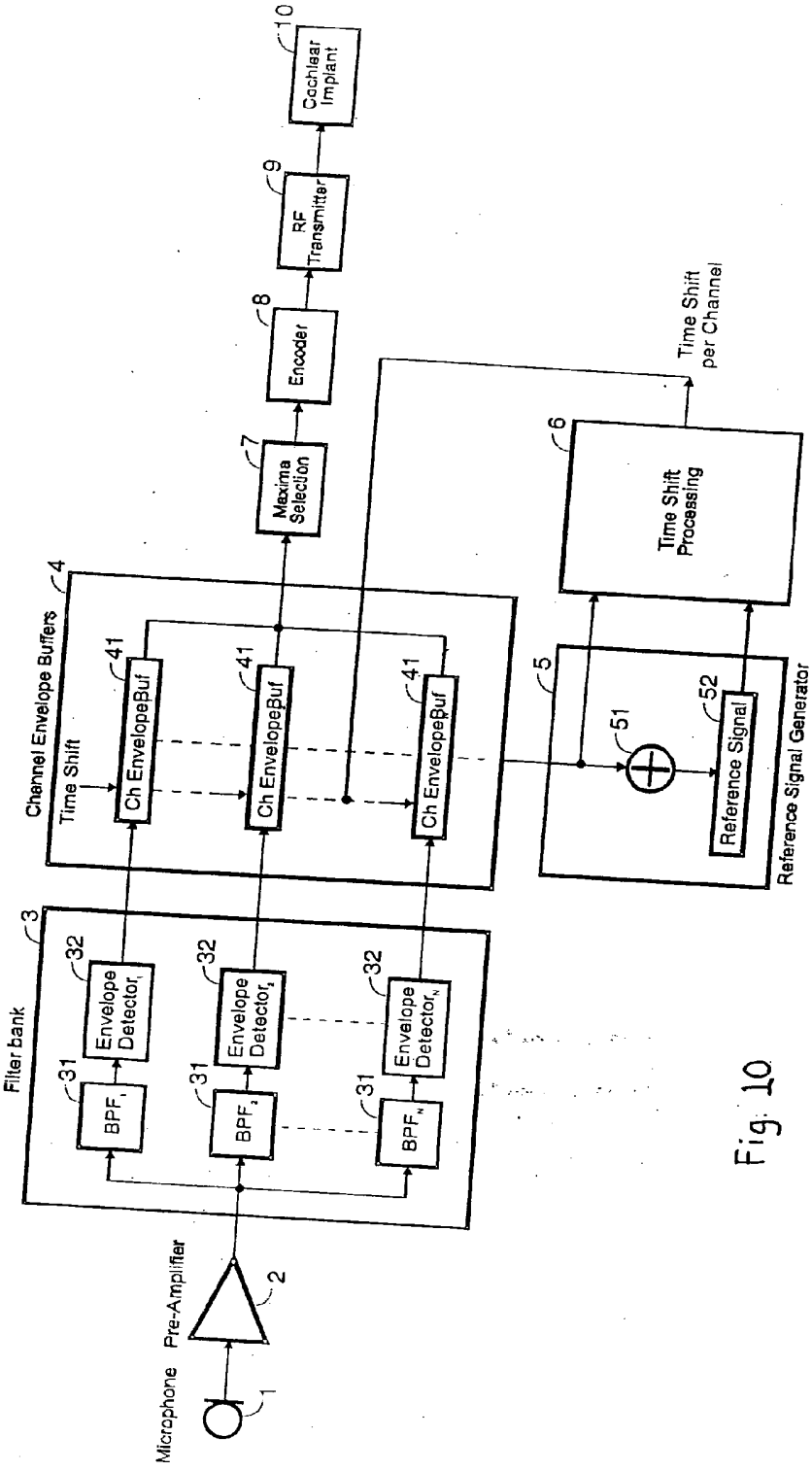
Microphone
1
Pre-Amplifier
2
Filter bank
3
BPF1
BPF2
BPFN
31
Envelope Detector1
Envelope Detector2
Envelope DetectorN
32
Channel Envelope Buffers
4
Time Shift
Ch EnvelopeBuf
41
Reference Signal Generator
5
51
Reference Signal
52
Time Shift Processing
6
Time Shift per Channel
Maxima Selection
7
Encoder
8
RF Transmitter
9
Cochlear Implant
10

Fig. 10

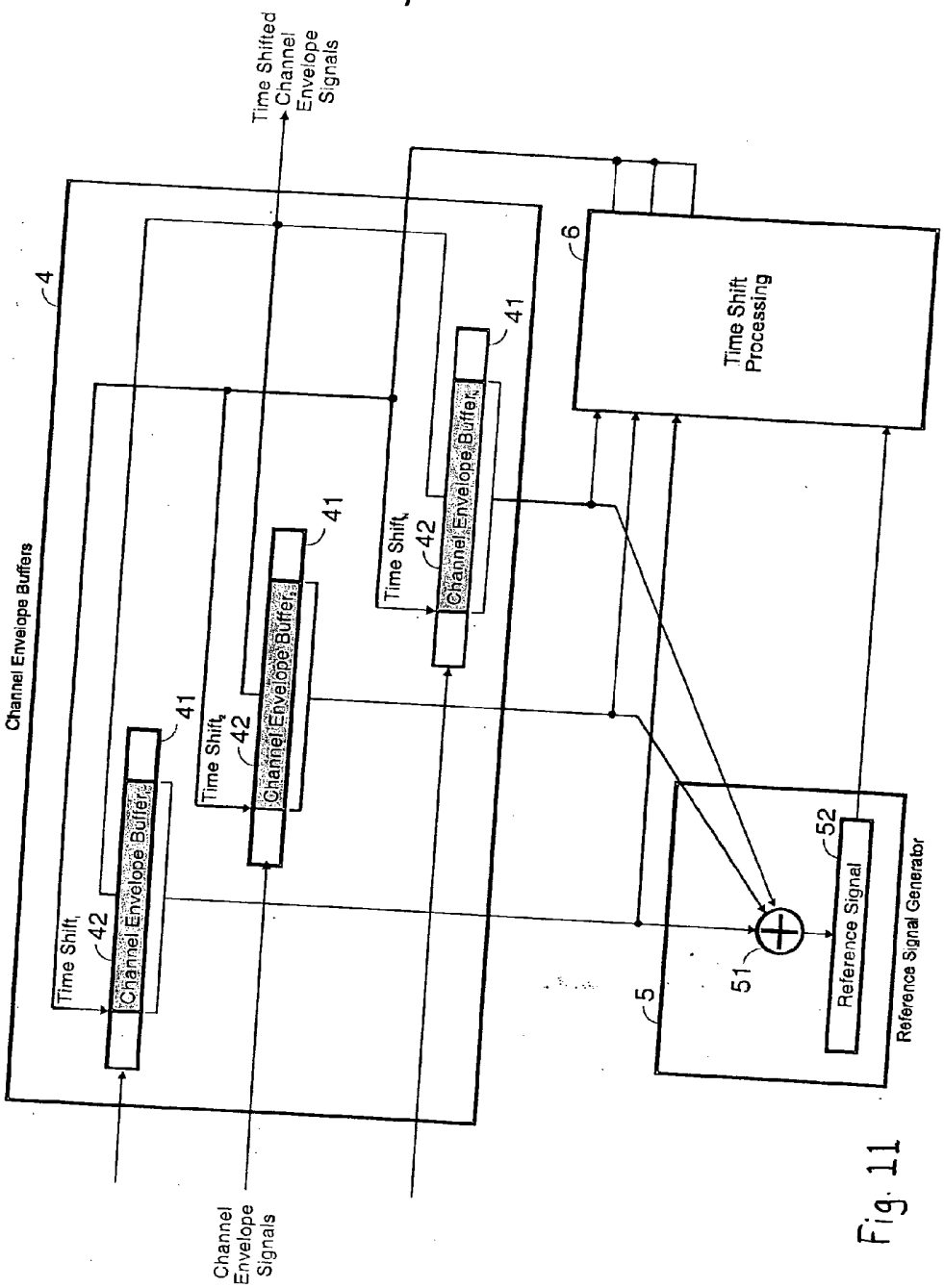
Time Shifted Channel Envelope Signals
Channel Envelope Buffers
4
41
Time Shift1
42
Channel Envelope Buffer1
Time Shift2
Channel Envelope Buffer2
Time ShiftN
Channel Envelope BufferN
6
Time Shift Processing
5
51
52
Reference Signal
Reference Signal Generator
Channel Envelope Signals

Fig. 11

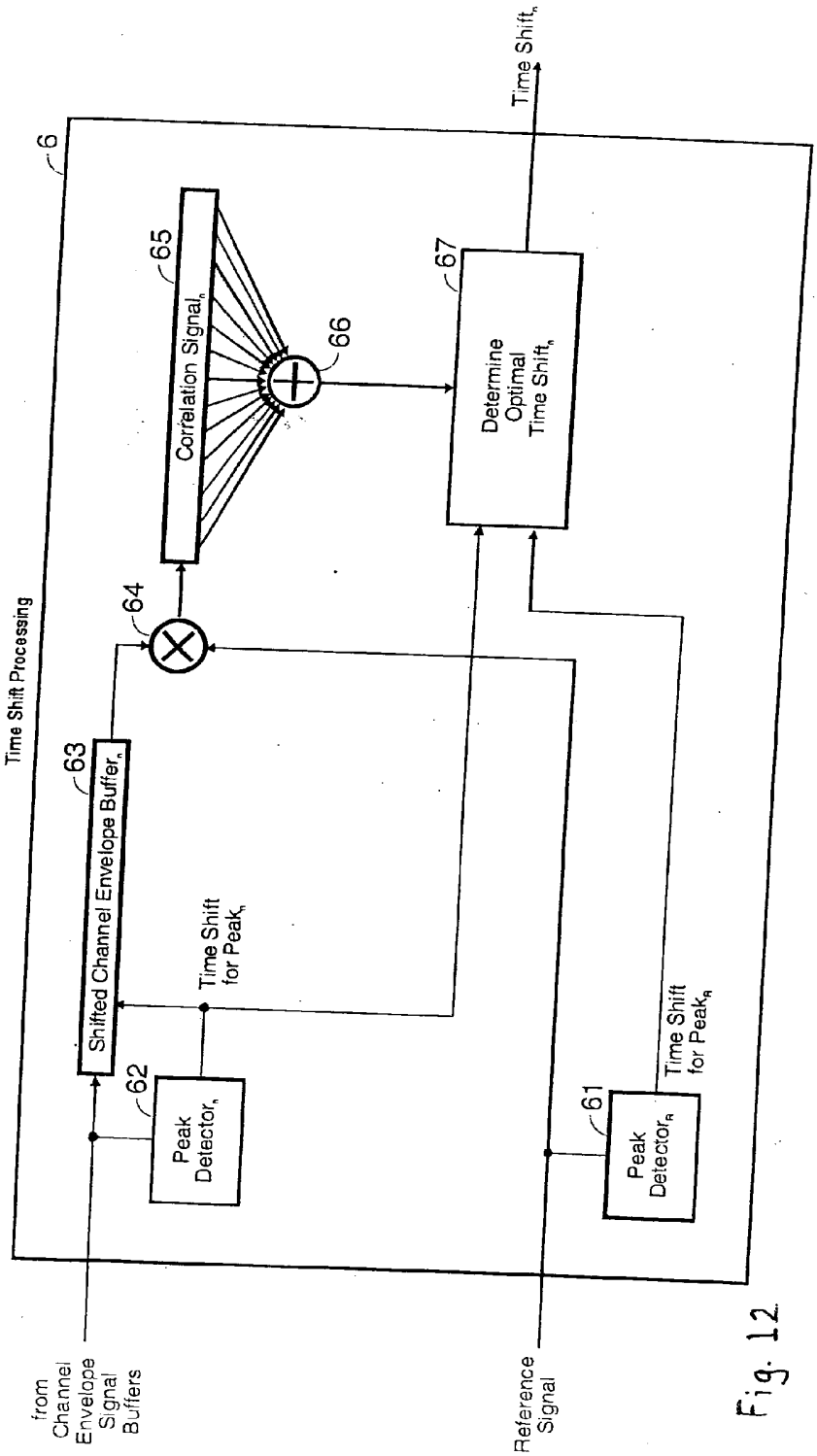
Time Shift Processing
6
Time Shift$_n$
65
Correlation Signal$_n$
66
67
Determine Optimal Time Shift$_n$
64
63
Shifted Channel Envelope Buffer$_n$
Time Shift for Peak$_n$
62
Peak Detector$_n$
61
Peak Detector$_R$
Time Shift for Peak$_R$
from Channel Envelope Signal Buffers
Reference Signal

Fig. 12