# PATROLL Winning Submission

## U.S. Patent 8,140,327

U.S. Patent 8,140,327 ("*Dialect LLC*" or the "patent-at-issue") was filed on April 22, 2010 and claims the benefit of U.S. Provisional Pat. App. 60/384,388, filed on June 3, 2002. Claim 1 of the patent-at-issue is generally directed to a method for filtering and eliminating noise from natural language utterances received at a microphone array that adds one or more nulls to a beam pattern steered to point in a direction associated with a user speaking the natural language utterance. The one or more nulls notch out point or limited area noise sources from an input speech signal corresponding to the natural language utterance. Environmental noise is compared to said input speech signal to set one or more parameters associated with an adaptive filter coupled to the microphone array. Said input speech signal is passed to the adaptive filter that uses band shaping and notch filtering to remove narrow-band noise from it according to said parameters. Cross-talk and environmentally-caused echoes are suppressed in said input speech signal using adaptive echo cancellation in the adaptive filter. The said input speech signal is then passed through the adaptive filter to a speech coder that uses adaptive lossy audio compression to remove momentary gaps from it and variable rate sampling to compress and digitize it. The speech coder optimizes the adaptive lossy audio compression and the variable rate sampling to only preserve components said input speech signal that will be input to a speech recognition engine. The digitized input speech signal is then transmitted from a buffer in the speech coder to the speech recognition engine at a rate that depends on the available bandwidth between the speech coder and the speech recognition engine.

A primary reference, European Patent 0694833 ("*AT&T*"), was filed on July 19, 1995 and was published on January 31, 1996. The patent generally relates to a human-computer interactive communications system, which facilitates the transfer of information to and from a human and a computer based system. Specifically, it describes an Intelligent Human Interface System (IHIS), which effectively utilizes automatic speech recognition technology, beam-steerable audio microphone technology, virtual reality imaging, and adaptive speakerphone technology.

A primary reference, U.S. Patent 9,491,544 ("*Solos Technology*"), was filed on February 18, 2014 and claims the benefit of U.S. Provisional Pat. App. 60/309,462, filed on August 1, 2001. The patent generally relates to frequency domain signal extraction. A reference signal containing mostly undesired audio and is substantially void of desired audio is received and decomposed into at least two reference frequency components. The at least two reference frequency components are filtered with at least two adaptive filters to form at least two filtered reference frequency components and recombined in an IFFT component, to produce a filtered reference signal. A delayed signal containing both desired and undesired audio is input to an adder. The filtered reference signal is subtracted from the delayed signal to form an output signal containing the desired audio, which is then decomposed into at least two frequency components. The filtering is adapted with the at least two frequency components and inhibited intermittently with the adaptive filters to prevent cancellation of the desired audio. Frequency sub-bands may be employed. An acoustic element with a cardioid beam pattern may be used to acquire the reference signal.

A secondary reference, U.S. Patent 5,715,319 ("*Polycom*"), was filed on May 30, 1996 and claim priority on the same date. The patent generally relates to steerable and endfire superdirective microphone arrays. Analog filters are used to band-limit at least two secondary microphone elements which are spaced from a primary microphone element a distance respective of their band limited outputs. The band-limited secondary microphone outputs are combined by an analog summer and the primary microphone and combined secondary microphone signals are digitized by an analog-to-digital converter. A signal processor performs a super-directive analysis of the primary microphone signal and the combined secondary microphone signals. The microphone outputs are digitized, split into frequency bands, and weighted sums are formed for each of a plurality of directions. A steering control circuit evaluates the relative energy of each directional signal in each band and selects a microphone direction for further processing and output.

A sample claim chart comparing claim 1 of *Dialect LLC* to *AT&T*, *Solos Technology*, and *Polycom* is provided below.

| US8140327 (“*Dialect LLC*”) | A. **EP0694833** (“*AT&T*”)<br>B. **US9491544** (“*Solos Technology*”)<br>C. **US5715319** (“*Polycom*”) |
|---|---|
| 1.pre. **A method for filtering and eliminating noise from natural language utterances**, comprising: | **A. EP0694833**<br>“**The microphone processor** 26 **uses the weighted signals to direct the spatial nulls of the microphone array beam toward sources of noise while directing the main lobe of the microphone array beam towards the desired source of audio**.” *AT&T* at p. 4:53-55<br><br>“**Regardless of the source towards which the beam is pointed, a spectrally accurate (clean) signal representation of audio data is received by the microphone processor** 26 **without corruption by the noise sources** 62 **and** 62a **and/or any other sources of extraneous interference**.” *AT&T* at p. 5:27-29<br><br>“**At the outset of the speech utterance by the customer** 41, **the IHIS** 50 **steers the microphone array** 34 **directly to the face of the customer** 41 **such that the speech of the customer** 41 **is received clearly while the noise sources** 62 **and** 62a **are not received. The speech data are processed through filter, gain and offset circuits of the microphone processor** 26 **and audio processor** 24 **before being input to the voice recognition engine** 20.” *AT&T* at p. 8:15-18<br><br>**B. US9491544**<br>“In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**.” *Solos Technology* at col. 3:30-33<br><br>“**Acoustic device** 102, **in accordance with the present invention, is designed to respond to audio presence, desired and undesired (i.e. noise), by outputting two audio beams** 103a **and** 103b, **with audio beam** 103a **having mostly undesired audio, substantially void of desired audio, and audio beam** 103b **having both the desired and the undesired audio**.” *Solos Technology* at col. 4:33-39<br><br>“**The two acoustic beams (hereinafter, simply beams) are sampled by signal processing subsystem** 104 **to generate two corresponding audio signals (hereinafter, simply** |

| | |
|---|---|
| (cont.)<br>1.pre. **A method for filtering and eliminating noise from natural language utterances**, comprising: | **signals), which in turn are used by signal processing subsystem** 104 **to recover the desired audio, by removing the first signal corresponding to the first beam from the second signal corresponding to the second beam**." *Solos Technology* at col. 4:40-46<br><br>"In one embodiment, **the echo cancellation like signal extraction process implemented is an adaptive noise cancellation process employing a NLMS FIR filter** (FIG. 11a). In other embodiments, **the echo cancellation like signal extraction processes implemented are adaptive noise cancellation processes employing a number of frequency domain LMS filters** (FIG. 11b-11c). In yet other embodiments, **the echo cancellation like signal extraction processes implemented are adaptive noise cancellation processes employing a number subband LMS filter** (FIG. 11d-11e)." *Solos Technology* at col. 9:42-51<br><br>**C. US5715319**<br>"**A steerable superdirective microphone array in accordance with another aspect of the present invention includes a first and a second microphone each having a forward directional response and a rearward directional response**." *Polycom* at col. 2:8-11<br><br>"**An analog-to-digital converter connected to receive signals from the first and second microphones produces digital signals representative of the microphone signals**." *Polycom* at col. 2:15-18<br><br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals**." *Polycom* at col. 2:35-39 |
| 1.a. **receiving a natural language utterance at a microphone array that adds one or more nulls to a beam pattern steered to point in a direction associated with a user speaking the natural language utterance**, | **A. EP0694833**<br>"**A first two-dimensional microphone array** 34 **is used to receive audible speech data from the customer** 41 **and/or companion** 41a. **The first microphone array is beam-steerable thus facilitating the reception of a non-corrupted audio signal from the audio source (customer** 41 **and/or companion** 41a). A second two-dimensional microphone array 34a is also beam-steerable and is used to scan the area in which the customer 41 is located in order to recognize additional sources of audio data (companions 41a) and sources of noise |

| | |
|---|---|
| (cont.)<br>1.a. **receiving a natural language utterance at a microphone array that adds one or more nulls to a beam pattern steered to point in a direction associated with a user speaking the natural language utterance,** | and/or interference 62, 62a." *AT&T* at p. 3:56-58 through p. 4:1-2<br><br>"**The microphone processor** 26 **uses the weighted signals to direct the spatial nulls of the microphone array beam toward sources of noise while directing the main lobe of the microphone array beam towards the desired source of audio**." *AT&T* at p. 4:53-55<br><br>"**When the companion** 41a **produces a speech utterance which reaches a pre-determined decibel level in excess of the customer** 41, **the central processor** 18 **provides a distinct set of delay values to the microphone processor** 26 **which re-directs the main beam of the first microphone array** 34 **towards the companion** 41a. **Likewise, when the customer** 41 **produces a speech utterance which reaches the predetermined decibel level in excess of the companion** 41a, **the main beam is again directed toward the customer** 41." *AT&T* at p. 5:19-23<br><br>"**At the outset of the speech utterance by the customer** 41, **the IHIS** 50 **steers the microphone array** 34 **directly to the face of the customer** 41 **such that the speech of the customer** 41 **is received clearly while the noise sources** 62 **and** 62a **are not received**." *AT&T* at p. 8:15-18<br><br>**B. US9491544**<br>"In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"**Acoustic device** 102, **in accordance with the present invention, is designed to respond to audio presence, desired and undesired (i.e. noise), by outputting two audio beams** 103a **and** 103b, **with audio beam** 103a **having mostly undesired audio, substantially void of desired audio, and audio beam** 103b **having both the desired and the undesired audio**." *Solos Technology* at col. 4:33-39<br><br>"**The two acoustic beams (hereinafter, simply beams) are sampled by signal processing subsystem** 104 **to generate two corresponding audio signals (hereinafter, simply signals), which in turn are used by signal processing subsystem** 104 **to recover the desired audio, by removing** |

| | |
|---|---|
| (cont.)<br>1.a. **receiving a natural language utterance at a microphone array that adds one or more nulls to a beam pattern steered to point in a direction associated with a user speaking the natural language utterance**, | **the first signal corresponding to the first beam from the second signal corresponding to the second beam**." *Solos Technology* at col. 4:40-46<br><br>"In each of these embodiments, one of the two cardioid beam generating mics 202a is arranged with its null facing the expected originating direction of desired audio, and **the other cardioid beam generating mic** 202b **arranged with its null facing away from the expected originating direction of desired audio**." *Solos Technology* at col. 5:60-65<br><br>"In summary, **acoustic device** 102 **comprises two or more acoustic elements designed and arranged in a manner that facilitates generation of two signals with one signal comprising mostly undesired audio, substantially void of desired audio, and another signal comprising both desired and undesired audio. The two or more acoustic elements may e.g. respond to the presence of audio, desired and undesired, outputting** a cardioid beam having a null facing the originating direction of desired audio, and **another beam having any one of a number of complementary beam shapes (as long as it does not comprise a null facing the originating direction of desired audio)**." *Solos Technology* at col. 6:63-67 through col. 7:1-7<br><br>**C. US5715319**<br>"**A directional microphone array in accordance with one aspect of the present invention includes a primary microphone connected to a first analog-to-digital converter and two or more secondary microphones arranged in line with and spaced predetermined distances from the primary microphone**." *Polycom* at col. 1:47-52<br><br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals**." *Polycom* at col. 2:35-39 |
| 1.b. **wherein the one or more nulls notch out point or limited area noise sources from an input speech signal corresponding to the natural language utterance**; | **A. EP0694833**<br>"**A first two-dimensional microphone array** 34 **is used to receive audible speech data from the customer** 41 **and/or companion** 41a. **The first microphone array is beam-steerable thus facilitating the reception of a non-corrupted audio signal from the audio source (customer** 41 **and/or companion** 41a). **A second two-dimensional microphone** |

| | |
|---|---|
| (cont.)<br>1.b. **wherein the one or more nulls notch out point or limited area noise sources from an input speech signal corresponding to the natural language utterance**; | **array** 34a **is also beam-steerable and is used to scan the area in which the customer** 41 **is located in order to recognize additional sources of audio data (companions** 41a) **and sources of noise and/or interference** 62, 62a." *AT&T* at p. 3:56-58 through p. 4:1-2<br><br>"**The microphone processor** 26 **uses the weighted signals to direct the spatial nulls of the microphone array beam toward sources of noise while directing the main lobe of the microphone array beam towards the desired source of audio**." *AT&T* at p. 4:53-55<br><br>"**Regardless of the source towards which the beam is pointed, a spectrally accurate (clean) signal representation of audio data is received by the microphone processor** 26 **without corruption by the noise sources** 62 **and** 62a **and/or any other sources of extraneous interference**." *AT&T* at p. 5:27-29<br><br>"**At the outset of the speech utterance by the customer** 41, **the IHIS** 50 **steers the microphone array** 34 **directly to the face of the customer** 41 **such that the speech of the customer** 41 **is received clearly while the noise sources** 62 **and** 62a **are not received**." *AT&T* at p. 8:15-18<br><br>**B. US9491544**<br>"In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"In each of these embodiments, one of the two cardioid beam generating mics 202a is arranged with its null facing the expected originating direction of desired audio, and **the other cardioid beam generating mic** 202b **arranged with its null facing away from the expected originating direction of desired audio**." *Solos Technology* at col. 5:60-65<br><br>"In summary, **acoustic device** 102 **comprises two or more acoustic elements designed and arranged in a manner that facilitates generation of two signals with one signal comprising mostly undesired audio, substantially void of desired audio, and another signal comprising both desired and undesired audio. The two or more acoustic elements may e.g. respond to the presence of audio, desired and** |

| | |
|---|---|
| (cont.)<br>1.b. **wherein the one or more nulls notch out point or limited area noise sources from an input speech signal corresponding to the natural language utterance**; | **undesired, outputting** a cardioid beam having a null facing the originating direction of desired audio, and **another beam having any one of a number of complementary beam shapes (as long as it does not comprise a null facing the originating direction of desired audio)**." *Solos Technology* at col. 6:63-67 through col. 7:1-7 |
| 1.c. **comparing environmental noise to the input speech signal corresponding to the natural language utterance to set one or more parameters associated with an adaptive filter coupled to the microphone array**; | **A. EP0694833**<br>"**At the outset of the speech utterance by the customer** 41, **the IHIS** 50 **steers the microphone array** 34 **directly to the face of the customer** 41 **such that the speech of the customer** 41 **is received clearly while the noise sources** 62 **and** 62a **are not received. The speech data are processed through filter, gain and offset circuits of the microphone processor** 26 **and audio processor** 24 **before being input to the voice recognition engine** 20." *AT&T* at p. 8:15-18<br><br>"Referring to FIGURE 2, **additional audio sources are identified as a companion** 41a, **and extraneous noise sources** 62 **and** 62a. **The central processor** 18 **identifies the extraneous noise sources** 62 **and** 62a **as such by calculating the short-time and long-time signal amplitude averages for each beam direction of the second microphone array** 34a. **Since human speech and extraneous noise have distinctly different relationships between their respective short-time and long-time amplitude averages, the central processor** 18 **differentiates between the two**." *AT&T* at p. 5:12-17<br><br>"**Thus the central processor** 18 **receives audio data from multiple sources dependant on the amplitude of the audible energy emitted therefrom. Regardless of the source towards which the beam is pointed, a spectrally accurate (clean) signal representation of audio data is received by the microphone processor** 26 **without corruption by the noise sources** 62 **and** 62a **and/or any other sources of extraneous interference**." *AT&T* at p. 5:26-29<br><br>"**The audio processor** 24, **adjusts the audio volume to a proper level to facilitate an improved conversation with the customer** 41. **The audio processor** 24 **increases the signal amplitude input to the loudspeaker** 32 **in accordance with commands from the central processor** 18." *AT&T* at p. 5:31-33<br><br>**B. US9491544**<br>"In one embodiment, **the desired audio is speech, in** |

| | |
|---|---|
| (cont.)<br>1.c. **comparing environmental noise to the input speech signal corresponding to the natural language utterance to set one or more parameters associated with an adaptive filter coupled to the microphone array**; | **particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"In summary, acoustic device 102 **comprises two or more acoustic elements designed and arranged in a manner that facilitates generation of two signals with one signal comprising mostly undesired audio, substantially void of desired audio, and another signal comprising both desired and undesired audio. The two or more acoustic elements may e.g. respond to the presence of audio, desired and undesired, outputting** a cardioid beam having a null facing the originating direction of desired audio, and **another beam having any one of a number of complementary beam shapes (as long as it does not comprise a null facing the originating direction of desired audio)**." *Solos Technology* at col. 6:63-67 through col. 7:1-7<br><br>"**The extraction logic operates as a loop running on a sample-by-sample basis. The reference signal is filtered by the adaptive FIR filter** 1102. **Essentially, a transfer function is applied to the reference channel to model the acoustic path from the cardioid element to the other element, so that the filtered reference signal closely matches the noise component of the signal in the primary channel. The filtered reference signal is then subtracted from the delayed primary signal. What is left, is the desired audio**." *Solos Technology* at col. 11:38-46<br><br>"**The (conditioned) signals of the reference channel and the delayed primary channel are first "decomposed" into a number of frequency components (two shown), by the corresponding FFT components** 1112. **Each of the frequency components of the reference signal is filtered by a corresponding adaptive filter** 1114, **and subtracted from the corresponding frequency component of the delayed signal of the primary channel, using a corresponding adder** 1116. **The resulted frequency components are "recombined", using IFFT component** 1118, **and the recombined signal is outputted as the desired audio**." *Solos Technology* at col. 11:57-67<br><br>"1. A system for signal extraction in a frequency domain, comprising:<br>. . . |

| | |
|---|---|
| (cont.)<br>1.c. **comparing environmental noise to the input speech signal corresponding to the natural language utterance to set one or more parameters associated with an adaptive filter coupled to the microphone array**; | **at least two adaptive filters, each of the at least two reference frequency components is filtered by a corresponding adaptive filter of the at least two adaptive filters to provide filtered reference frequency components; at least two adders, each of the at least two primary frequency components is input into a corresponding adder of the at least two adders, wherein a corresponding filtered reference frequency component is subtracted from a corresponding primary frequency component in the corresponding adder, an output from the corresponding adder is fed back to the corresponding adaptive filter to provide information that is used during the filtering at the corresponding adaptive filter**;" *Solos Technology* at claim 1<br><br>**C. US5715319**<br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals**." *Polycom* at col. 2:35-39<br><br>"**After all the bands are tallied, the direction which received the greatest number of votes is selected for output during the current sample provided that the number of votes is greater than a predetermined minimum, for example, seven, indicating that the signal is significantly stronger than the noise**." *Polycom* at col. 9:6-11<br><br>"1. **A directional microphone array comprising**:<br>. . .<br>**an analog frequency filter connected to said secondary microphones for respectively limiting said output of each of said secondary microphones to a predetermined frequency band having a predetermined relationship to said respective offset and providing frequency filtered outputs respective of said secondary microphones**;" *Polycom* at claim 1 |
| 1.d. **passing the input speech signal corresponding to the natural language utterance to the adaptive filter**, | **A. EP0694833**<br>"**At the outset of the speech utterance by the customer** 41, **the IHIS** 50 **steers the microphone array** 34 **directly to the face of the customer** 41 **such that the speech of the customer** 41 **is received clearly while the noise sources** 62 **and** 62a **are not received. The speech data are processed through filter, gain and offset circuits of the microphone processor** 26 **and audio processor** 24 **before being input to** |

| | |
|---|---|
| (cont.)<br>1.d. **passing the input speech signal corresponding to the natural language utterance to the adaptive filter**, | **the voice recognition engine** 20." *AT&T* at p. 8:15-18<br><br>**B. US9491544**<br>"In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"**The extraction logic operates as a loop running on a sample-by-sample basis. The reference signal is filtered by the adaptive FIR filter** 1102. **Essentially, a transfer function is applied to the reference channel to model the acoustic path from the cardioid element to the other element, so that the filtered reference signal closely matches the noise component of the signal in the primary channel. The filtered reference signal is then subtracted from the delayed primary signal. What is left, is the desired audio**." *Solos Technology* at col. 11:38-46<br><br>"**The (conditioned) signals of the reference channel and the delayed primary channel are first "decomposed" into a number of frequency components (two shown), by the corresponding FFT components** 1112. **Each of the frequency components of the reference signal is filtered by a corresponding adaptive filter** 1114, **and subtracted from the corresponding frequency component of the delayed signal of the primary channel, using a corresponding adder** 1116. **The resulted frequency components are "recombined", using IFFT component** 1118, **and the recombined signal is outputted as the desired audio**." *Solos Technology* at col. 11:57-67<br><br>"1. A system for signal extraction in a frequency domain, comprising:<br>. . .<br>**at least two adaptive filters, each of the at least two reference frequency components is filtered by a corresponding adaptive filter of the at least two adaptive filters to provide filtered reference frequency components; at least two adders, each of the at least two primary frequency components is input into a corresponding adder of the at least two adders, wherein a corresponding filtered reference frequency component is subtracted from a corresponding primary frequency component in the corresponding adder, an output from the corresponding** |

| | |
|---|---|
| (cont.)<br>1.d. **passing the input speech signal corresponding to the natural language utterance to the adaptive filter**, | **adder is fed back to the corresponding adaptive filter to provide information that is used during the filtering at the corresponding adaptive filter**;" *Solos Technology* at claim 1<br><br>**C. US5715319**<br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals**." *Polycom* at col. 2:35-39<br><br>"**A frequency filter connected to the input receives the microphone signals and produces a plurality of narrow band signals respective of each one of the microphones as an output**." *Polycom* at col. 3:67 through col. 4:1-3 |
| 1.e. **wherein the adaptive filter uses band shaping and notch filtering to remove narrow-band noise from the input speech signal corresponding to the natural language utterance according to the one or more parameters**; | **B. US9491544**<br>"In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"**The signal from each acoustic element is amplified by a corresponding pre-amp** 602, **and then band-limited by a corresponding anti-aliasing filter** 604, before being digitized by a corresponding A/D converter at the sampling frequency Fs." *Solos Technology* at col. 7:50-54<br><br>"**For these applications, a pre-whitening filter (also referred to as de-colorization filter) is placed on both the primary and reference inputs before they are sent to signal extraction component** 506, **in particular, if component** 506 **implements NMLS noise cancellation processing, to alleviate the potential slow convergence rate brought about by narrow band (highly auto-correlated) input signal**." *Solos Technology* at col. 8:10-16<br><br>"As alluded to earlier, described in more detail below, **filtering may also be performed in the** frequency and **subband domains**." *Solos Technology* at col. 11:16-18<br><br>"**The extraction logic operates as a loop running on a sample-by-sample basis. The reference signal is filtered by the adaptive FIR filter** 1102. **Essentially, a transfer function is applied to the reference channel to model the** |

| | |
|---|---|
| (cont.)<br>1.e. **wherein the adaptive filter uses band shaping and notch filtering to remove narrow-band noise from the input speech signal corresponding to the natural language utterance according to the one or more parameters**; | **acoustic path from the cardioid element to the other element, so that the filtered reference signal closely matches the noise component of the signal in the primary channel. The filtered reference signal is then subtracted from the delayed primary signal. What is left, is the desired audio**." *Solos Technology* at col. 11:38-46<br><br>"**The (conditioned) signals of the reference channel and the delayed primary channel are first "decomposed" into a number of frequency components (two shown), by the corresponding FFT components** 1112. **Each of the frequency components of the reference signal is filtered by a corresponding adaptive filter** 1114, **and subtracted from the corresponding frequency component of the delayed signal of the primary channel, using a corresponding adder** 1116. **The resulted frequency components are "recombined", using IFFT component** 1118, **and the recombined signal is outputted as the desired audio**." *Solos Technology* at col. 11:57-67<br><br>"1. A system for signal extraction in a frequency domain, comprising:<br>. . .<br>**at least two adaptive filters, each of the at least two reference frequency components is filtered by a corresponding adaptive filter of the at least two adaptive filters to provide filtered reference frequency components; at least two adders, each of the at least two primary frequency components is input into a corresponding adder of the at least two adders, wherein a corresponding filtered reference frequency component is subtracted from a corresponding primary frequency component in the corresponding adder, an output from the corresponding adder is fed back to the corresponding adaptive filter to provide information that is used during the filtering at the corresponding adaptive filter**;" *Solos Technology* at claim 1<br><br>**C. US5715319**<br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals**." *Polycom* at col. 2:35-39<br><br>"**A frequency filter connected to the input receives the microphone signals and produces a plurality of narrow** |

| | |
|---|---|
| (cont.)<br>1.e. **wherein the adaptive filter uses band shaping and notch filtering to remove narrow-band noise from the input speech signal corresponding to the natural language utterance according to the one or more parameters**; | **band signals respective of each one of the microphones as an output**." *Polycom* at col. 3:67 through col. 4:1-3<br><br>"**The optimized narrow band signal for each frequency band is synthesized into time domain signals and bandpass filtered, and then combined by a summer** 350 **to form the microphone array output**." *Polycom* at col. 6:64-67 through col. 7:1<br><br>"Alternatively, **various signal enhancement processes may be incorporated in the signal processor. For example, echo cancellation, noise suppression, automatic gain control, and speech compression may be performed on the optimized narrow band signals before the inverse FFT is performed thereby avoiding the added computational requirements and delay of a second bandpass analysis**." *Polycom* at col. 7:4-10<br><br>"**After all the bands are tallied, the direction which received the greatest number of votes is selected for output during the current sample provided that the number of votes is greater than a predetermined minimum, for example, seven, indicating that the signal is significantly stronger than the noise**." *Polycom* at col. 9:6-11<br><br>"**Using the combined outputs of two rings of band-limited microphones provides an enhanced signal-to-noise ratio in the superdirective array because the apparent spacing of the real and virtual elements in the array relative to each other increases with decreasing frequency**." *Polycom* at col. 9:54-59<br><br>"20. The method of claim 19 further comprising the steps of: **performing at least one process for echo cancellation, noise suppression, automatic gain control, or speech compression using said selected ones of said narrow band directional signals**." *Polycom* at claim 20 |
| 1.f. **suppressing cross-talk and environmentally caused echoes in the input speech signal corresponding to the natural language utterance using adaptive echo cancellation in the adaptive filter**; | **A. EP0694833**<br>"**A first two-dimensional microphone array** 34 **is used to receive audible speech data from the customer** 41 **and/or companion** 41a. **The first microphone array is beam-steerable thus facilitating the reception of a non-corrupted audio signal from the audio source (customer** 41 **and/or companion** 41a). **A second two-dimensional microphone array** 34a **is also beam-steerable and is used to scan the** |

| | |
|---|---|
| (cont.)<br>1.f. **suppressing cross-talk and environmentally caused echoes in the input speech signal corresponding to the natural language utterance using adaptive echo cancellation in the adaptive filter**; | **area in which the customer** 41 **is located in order to recognize additional sources of audio data (companions** 41a) **and sources of noise and/or interference** 62, 62a." *AT&T* at p. 3:56-58 through p. 4:1-2<br><br>"Referring to FIGURE 2, **additional audio sources are identified as a companion** 41a, **and extraneous noise sources** 62 **and** 62a. **The central processor** 18 **identifies the extraneous noise sources** 62 **and** 62a **as such by calculating the short-time and long-time signal amplitude averages for each beam direction of the second microphone array** 34a. **Since human speech and extraneous noise have distinctly different relationships between their respective short-time and long-time amplitude averages, the central processor** 18 **differentiates between the two. In a like manner, the central processor** 18 **identifies a bona fide additional source of audio data as that from the companion** 41a." *AT&T* at p. 5:12-18<br><br>"**Regardless of the source towards which the beam is pointed, a spectrally accurate (clean) signal representation of audio data is received by the microphone processor** 26 **without corruption by the noise sources** 62 **and** 62a **and/or any other sources of extraneous interference**." *AT&T* at p. 5:26-29<br><br>"In an alternate embodiment of the present invention, **the central processor** 18, **audio processor** 24 **and microphone processor** 26 **work in synergy to achieve the function of an acoustic echo canceling system as defined in United States Patent No. 5,001,701. An acoustic echo canceling system also achieves near full duplex performance while eliminating the aforementioned feedback problem**." *AT&T* at p. 6:18-21<br><br>"**When an acoustic echo canceling system is used, the central processor** 18 **generates a signal which represents the spectral response of the channel between the loudspeaker** 32 **and the microphone array** 34 **to a signal to be projected from the loudspeaker** 32. **The central processor** 18 **then combines the generated signal with the signal to be projected from the loudspeaker** 32 **to cancel any energy which would ordinarily feed back to the microphone array** 34." *AT&T* at p. 6:22-25<br><br>"**At the outset of the speech utterance by the customer** 41, |

| | |
|---|---|
| (cont.)<br>1.f. **suppressing cross-talk and environmentally caused echoes in the input speech signal corresponding to the natural language utterance using adaptive echo cancellation in the adaptive filter**; | **the IHIS** 50 **steers the microphone array** 34 **directly to the face of the customer** 41 **such that the speech of the customer** 41 **is received clearly while the noise sources** 62 **and** 62a **are not received. The speech data are processed through filter, gain and offset circuits of the microphone processor** 26 **and audio processor** 24 **before being input to the voice recognition engine** 20." *AT&T* at p. 8:15-18<br><br>**B. US9491544**<br>"In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"In one embodiment, **the echo cancellation like signal extraction process implemented is an adaptive noise cancellation process employing a NLMS FIR filter** (FIG. 11a). In other embodiments, **the echo cancellation like signal extraction processes implemented are adaptive noise cancellation processes employing a number of frequency domain LMS filters** (FIG. 11b-11c). In yet other embodiments, **the echo cancellation like signal extraction processes implemented are adaptive noise cancellation processes employing a number subband LMS filter** (FIG. 11d-11e)." *Solos Technology* at col. 9:42-51<br><br>"**The extraction logic operates as a loop running on a sample-by-sample basis. The reference signal is filtered by the adaptive FIR filter** 1102. **Essentially, a transfer function is applied to the reference channel to model the acoustic path from the cardioid element to the other element, so that the filtered reference signal closely matches the noise component of the signal in the primary channel. The filtered reference signal is then subtracted from the delayed primary signal. What is left, is the desired audio**." *Solos Technology* at col. 11:38-46<br><br>"**The (conditioned) signals of the reference channel and the delayed primary channel are first "decomposed" into a number of frequency components (two shown), by the corresponding FFT components** 1112. **Each of the frequency components of the reference signal is filtered by a corresponding adaptive filter** 1114, **and subtracted from the corresponding frequency component of the delayed signal of the primary channel, using a corresponding adder** |

| | |
|---|---|
| (cont.)<br>1.f. **suppressing cross-talk and environmentally caused echoes in the input speech signal corresponding to the natural language utterance using adaptive echo cancellation in the adaptive filter**; | 1116. **The resulted frequency components are "recombined", using IFFT component** 1118, **and the recombined signal is outputted as the desired audio.**" *Solos Technology* at col. 11:57-67<br><br>**C. US5715319**<br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals.**" *Polycom* at col. 2:35-39<br><br>"**A frequency filter connected to the input receives the microphone signals and produces a plurality of narrow band signals respective of each one of the microphones as an output.**" *Polycom* at col. 3:67 through col. 4:1-3<br><br>"**Echo cancellation is disclosed in U.S. Pat. No. 5,305,307 entitled "Adaptive Acoustic Echo Canceller Having Means for Reducing or Eliminating Echo in a Plurality of Signal Bandwidths" and in U.S. Pat. No. 5,263,019, entitled "Method and Apparatus for Estimating the Level of Acoustic Feedback Between a Loudspeaker and Microphone"; noise suppression is disclosed in copending application Ser. No. 08/402,550, entitled "Reduction Of Background Noise for Speech Enhancement", filed on Mar. 13, 1995**; . . . ." *Polycom* at col. 7:10-19 |
| 1.g. **sending the input speech signal passed through the adaptive filter to a speech coder that uses adaptive lossy audio compression to remove momentary gaps from the input speech signal and variable rate sampling to compress and digitize the input speech signal**, | **B. US9491544**<br>"**Sampling components** 504 **are employed to digitized beams** 103a **and** 103b. **Typically, they are both digitized synchronically at the same sampling frequency, which is application dependent, and chosen according to the system bandwidth. In the case of ASR applications, the sampling frequency e.g. may be 8 kHz, 11 kHz, 12 kHz, or 16 kHz.**" *Solos Technology* at col. 7:21-26<br><br>"**The signal from each acoustic element is amplified by a corresponding pre-amp** 602, **and then band-limited by a corresponding anti-aliasing filter** 604, **before being digitized by a corresponding A/D converter at the sampling frequency Fs.**" *Solos Technology* at col. 7:50-54<br><br>"However, if both channels are active, **inhibit logic** 808 **further determines if either desired audio is present or a pause threshold (also referred to as hangover time) has not** |

| | |
|---|---|
| (cont.)<br>1.g. **sending the input speech signal passed through the adaptive filter to a speech coder that uses adaptive lossy audio compression to remove momentary gaps from the input speech signal and variable rate sampling to compress and digitize the input speech signal,** | **been reached**, block 1206-1208. **The pause threshold (or hangover time) is application dependent. For example, in the case of ASR, the pause threshold may be a fraction of a second**." *Solos Technology* at col. 10:61-67<br><br>"**If the desired audio is detected or the pause time is not exceeded, the inhibit signal is set to "positive with filter adaptation disabled", i.e. filtering coefficients frozen**, block 1212. **The reference signal is filtered accordingly, and subtracted from the primary channel to generate desired audio**." *Solos Technology* at col. 11:1-5<br><br>**C. US5715319**<br>"**A directional microphone array in accordance with one aspect of the present invention includes a primary microphone connected to a first analog-to-digital converter and two or more secondary microphones arranged in line with and spaced predetermined distances from the primary microphone. The two or more secondary microphones are each frequency filtered with the response of each secondary microphone being limited to a predetermined band of frequencies respective of the relative placement of the respective secondary microphone. The frequency filtered secondary microphone outputs are combined and input to a second analog-to-digital converter**." *Polycom* at col. 1:47-58<br><br>"**An analog-to-digital converter connected to receive signals from the first and second microphones produces digital signals representative of the microphone signals**." *Polycom* at col. 2:15-18<br><br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals**." *Polycom* at col. 2:35-39<br><br>"**Preferably, a Fast Fourier Transform is used to perform the narrow band analysis of filters** 310. **In a preferred embodiment, a 512 point FFT is performed on a group of 512 samples from each A/D channel thereby splitting each full band signal into 256 frequency bands. The A/D** 120 of FIG. 1 **may be operated at a sample rate of 16 KHz yielding 256 frequency bands of 31.25 Hz width in the range of 0 to 8 KHz. When 2× oversampling is used, an FFT is** |

| | |
|---|---|
| (cont.)<br>1.g. **sending the input speech signal passed through the adaptive filter to a speech coder that uses adaptive lossy audio compression to remove momentary gaps from the input speech signal and variable rate sampling to compress and digitize the input speech signal,** | **performed every 16 milliseconds for each channel.**" *Polycom* at col. 5:65-67 through col. 6:1-6 |
| 1.h. **wherein the speech coder optimizes the adaptive lossy audio compression and the variable rate sampling to only preserve components in the input speech signal that will be input to a speech recognition engine**; and | **B. US9491544**<br>"In one embodiment, **the desired audio is speech, in particular, speech to be recognized in a noisy reverberant environment, where noise is stochastic and uncorrelated to the desired speech, such as in an automobile or in an office**." *Solos Technology* at col. 3:30-33<br><br>"In the following description, **various embodiments of the present invention will be described, in particular, ASR oriented embodiments**." *Solos Technology* at col. 3:38-40<br><br>"**Part of the descriptions will employ various abbreviations, including but are not limited to**:<br>**ASR Automatic Speech Recognition**" *Solos Technology* at col. 4:1-3<br><br>"**The signal from each acoustic element is amplified by a corresponding pre-amp** 602, **and then band-limited by a corresponding anti-aliasing filter** 604, **before being digitized by a corresponding A/D converter at the sampling frequency Fs**." *Solos Technology* at col. 7:50-54<br><br>"**Sampling components** 504 **are employed to digitized beams** 103a **and** 103b. **Typically, they are both digitized synchronically at the same sampling frequency, which is application dependent, and chosen according to the system bandwidth. In the case of ASR applications, the sampling frequency e.g. may be 8 kHz, 11 kHz, 12 kHz, or 16 kHz**." *Solos Technology* at col. 7:21-26<br><br>"**The (conditioned) signals of the reference channel and the delayed primary channel are first "decomposed" into a number of frequency components (two shown), by the corresponding FFT components** 1112. **Each of the frequency components of the reference signal is filtered by a corresponding adaptive filter** 1114, **and subtracted from** |

| | |
|---|---|
| (cont.)<br>1.h. **wherein the speech coder optimizes the adaptive lossy audio compression and the variable rate sampling to only preserve components in the input speech signal that will be input to a speech recognition engine**; and | **the corresponding frequency component of the delayed signal of the primary channel, using a corresponding adder** 1116. **The resulted frequency components are "recombined", using IFFT component** 1118, **and the recombined signal is outputted as the desired audio.**" *Solos Technology* at col. 11:57-67<br><br>**C. US5715319**<br>"**An analog-to-digital converter connected to receive signals from the first and second microphones produces digital signals representative of the microphone signals.**" *Polycom* at col. 2:15-18<br><br>"**The steerable array may also have a signal processor connected to receive the signals in each band from the selected direction and perform echo cancellation, noise suppression, automatic gain control, or speech compression on the selected signals.**" *Polycom* at col. 2:35-39<br><br>"**Preferably, a Fast Fourier Transform is used to perform the narrow band analysis of filters** 310. **In a preferred embodiment, a 512 point FFT is performed on a group of 512 samples from each A/D channel thereby splitting each full band signal into 256 frequency bands. The A/D** 120 of FIG. 1 **may be operated at a sample rate of 16 KHz yielding 256 frequency bands of 31.25 Hz width in the range of 0 to 8 KHz. When 2× oversampling is used, an FFT is performed every 16 milliseconds for each channel.**" *Polycom* at col. 5:65-67 through col. 6:1-6 |
| 1.i. **transmitting the digitized input speech signal from a buffer in the speech coder to the speech recognition engine**, | **A. EP0694833**<br>"**An audio processor** 24 **suited for the requirements of the present invention is the S201 available from GBCS, a division of the AT&T Corporation.**" *AT&T* at p. 5:30-31<br><br>"**The 'clean' audio data is input to the voice recognition engine** 20 **from the audio processor** 24. Referring to FIGURE 5, **the voice recognition engine** 20 **and the voice synthesis engine** 22 **generally comprise a speech recognizer** 70, **language analyzer** 72, **expert system** 74 **and text to speech synthesizer** 22. **The voice recognition engine** 20 **is implemented using a combination of Digital Signal Processing (DSP) integrated circuits working in conjunction with the microprocessor controller** 18." *AT&T* at p. 6:26-30 |

| | |
|---|---|
| (cont.)<br>1.i. **transmitting the digitized input speech signal from a buffer in the speech coder to the speech recognition engine**, | “**During the conversation, the IHIS** 50 **stores a sample of the speech data of the customer in the memory** 16.” *AT&T* at p. 7:57-58<br><br>“At the outset of the speech utterance by the customer 41, the IHIS 50 steers the microphone array 34 directly to the face of the customer 41 such that the speech of the customer 41 is received clearly while the noise sources 62 and 62a are not received. **The speech data are processed through filter, gain and offset circuits of the microphone processor** 26 **and audio processor** 24 **before being input to the voice recognition engine** 20.” *AT&T* at p. 8:15-18<br><br>“**A Fourier Transform algorithm program stored in the memory** 16 **and executed by the central processor** 18 **produces the frequency spectrum of at least a portion of the speech data prior to being input to the speech recognizer** 70.” *AT&T* at p. 8:25-26 |
| 1.j. **wherein the speech coder transmits the digitized input speech signal to the speech recognition engine at a rate that depends on available bandwidth between the speech coder and the speech recognition engine**. | **B. US9491544**<br>“**Sampling components** 504 **are employed to digitized beams** 103a **and** 103b. **Typically, they are both digitized synchronically at the same sampling frequency, which is application dependent, and chosen according to the system bandwidth. In the case of ASR applications, the sampling frequency e.g. may be 8 kHz, 11 kHz, 12 kHz, or 16 kHz**.” *Solos Technology* at col. 7:21-26 |