(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
**23 November 2017 (23.11.2017)**

WIPO | PCT

(10) **International Publication Number**
**WO 2017/198737 A1**

(51) **International Patent Classification:**
*G11B 27/28* (2006.01) *G11B 27/00* (2006.01)
*G10L 19/02* (2013.01) *G10L 21/04* (2013.01)

(21) **International Application Number:** PCT/EP2017/061895

(22) **International Filing Date:** 17 May 2017 (17.05.2017)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
16170723.7 20 May 2016 (20.05.2016) EP
16179531.5 14 July 2016 (14.07.2016) EP

(71) **Applicant: FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastraße 27c, 80686 München (DE).

(72) **Inventors: UHLE, Christian**; Hoher Rain 28, 92289 Ursensollen (DE). **KRATZ, Michael**; Friedrichstraße 9, 91054 Erlangen (DE). **KLOSE, Paul**; Fritz-Erler-Straße 1, 90765 Fürth (DE). **LEONARD, Timothy**; Bauerngasse 19, 90443 Nürnberg (DE). **LUVIZOTTO, André**; Egidienplatz 1, 90403 Nürnberg (DE). **SCHARRER, Sebastian**; Beegstraße 26, 90518 Altdorf (DE).

(74) **Agent: ZINKLER, Franz** et al.; Schoppe, Zimmermann, Stöckeler, Zinkler, Schenk & Partner mbB, Radlkoferstr. 2, 81373 München (DE).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** APPARATUS AND METHOD FOR PROCESSING A MULTICHANNEL AUDIO SIGNAL

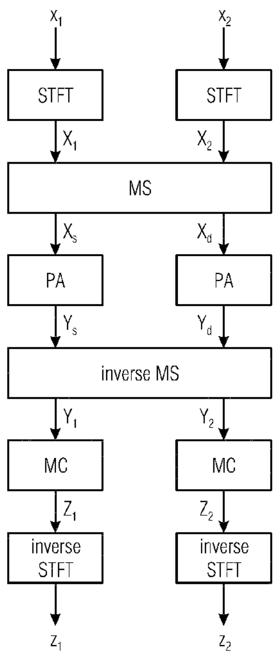


Fig. 2

**(57) Abstract:** An apparatus (1) for processing a multichannel audio signal (100) is provided, comprising a plurality of channel signals ($x_1$, $x_2$). The apparatus performs a time scale modulation of the multichannel audio signal (100) and comprises a phase adaptor (5) and a separator (6). The phase adaptor (5) provides a processed signal ($Y_s$, $Y_d$) by modifying a phase of a signal ($X_s$, $X_d$) based on a combination of the channel signals ($x_1$, $x_2$). The separator (6) provides separated signals ($Y_1$, $Y_2$) based on the processed signal ($Y_s$, $Y_d$). Also a corresponding method is provided.

WO 2017/198737 A1

*[Continued on next page]*

**Published:**

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

# Apparatus and method for processing a multichannel audio signal

## Specification

The invention refers to an apparatus for processing a multichannel audio signal. The multichannel audio signal comprises a plurality of – i.e. at least two – channel signals. The apparatus performs a time scale modulation of the multichannel audio signal. The invention also refers to a corresponding method and a computer program.

Time scale modification (TSM) refers to the processing to slow down or speed up the playback of an audio signal without affecting its pitch. TSM in combination with sample rate conversion also enables to change the pitch without modifying the tempo. The challenge for TSM is to maintain all other characteristics of the audio signal (except either tempo or pitch) and in particular the sound quality. The processing should not produce audible artefacts.

For single-channel input signals, the main important characteristic is the timbre. For signals having more than one channel, also spatial characteristics need to be maintained. The spatial characteristics comprise the position and width of the direct sound source and the diffuseness of ambient sound. They can be quantified by inter-channel level differences (ICLD), inter-channel time differences (ICTD), inter-channel phase differences (ICPD) or and inter-channel coherence (ICC).

Two fundamentally different approaches to time scale modification exist. One is applied in the time domain and the other is applied in the frequency domain.

Processing in the time domain uses a synchronised overlap-add (SOLA) scheme. The signal is cut into overlapping frames and these frames are shifted and combined to stretch or shrink the signal. The shifting position is typically computed by maximising a measure of similarity, e.g. correlation between the signal frame and its shifted copy.

This method in the time domain is of low computational complexity. It yields good results for monophonic (in contrast to polyphonic) signals, e.g. speech or flute tones, because the shifting offset can be determined as an integer multiple of the period of the fundamental frequency in order to avoid discontinuities and destructive interference in the output signal. In other words, the shifted signal frames are added in a phase coherent way. For polyphonic inputs with multiple tones having different fundamental frequencies, the shifting cannot be determined such that wave form similarity is fulfilled for all tones (pitches).

For many musical signals, better results in terms of sound quality are obtained by applying the processing in the frequency domain. This method uses e.g. the phase vocoder scheme [1] as shown in the block diagram shown in Fig. 1 and briefly explained in the following.

The input audio signal x(n) is transformed into the frequency domain using a short-time Fourier transform (STFT). Equivalently, another type of filterbank or transform can be used for which the inverse processing can be applied with sufficiently small reconstruction error.

In the mentioned embodiment, the input signal x(n) is cut into overlapping frames and a Discrete Fourier transform (DFT) is computed for each frame according to the following Equation (1), yielding a short-time Fourier transform (STFT) representation of the signal, also referred to as STFT coefficients (or spectral coefficients), according to

$$X(m,k) = \sum_{n=0}^{N-1} x(mR_a + n) w_a(n) \mathrm{e}^{-j\Omega_k n} \qquad (1)$$

A time frame index is denoted by m, k is a discrete frequency index with $0 \leq k \leq N-1$, and $w_a$ is a window function. The normalized angular frequency $\Omega_k$ is given by $\Omega_k = 2\pi k / N$. The DFT has size N and $R_a$ is an analysis hop size.

The indices for time and frequency are omitted in the description when possible for brevity.

The output time-domain signal y(n) is computed from output spectral coefficients Y(m, k) in the synthesis stage by means of the inverse of the STFT which is performed in two steps:

Firstly, an inverse Discrete Fourier transform is computed for each of the M frames according to

$$y_m(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y(m,k) \mathrm{e}^{j\Omega_k n} \tag{2}$$

Secondly, an overlap-add procedure is applied according to

$$y(n) = \sum_{m=0}^{M} w_s(n - mR_s) y_m(n - mR_s) \tag{3}$$

with optional synthesis window $w_s(n)$ and synthesis hop size $R_s$.

A time scale modification is achieved by setting the synthesis hop size $R_s$ and the analysis hop size $R_a$ to different values: the signal is stretched in time if $R_a < R_s$ and is shrinked if $R_a > R_s$.

The analysis window $w_a$ and synthesis window $w_s$ are chosen such that if $R_a = R_s$ and $Y(m, k) = X(m, k)$, then the input and output signal are identical.

Short-time Fourier transform coefficients $X(m, k)$ for real-valued input signals (which is the case for audio signals considered here) are complex-valued numbers that can be expressed in polar coordinates by their magnitude $|X|$ and phase $\Phi_x$ as

$$X = |X| \exp(j\Phi_x), \tag{4}$$

$$\Phi_x = \arg X, \tag{5}$$

where $j = \sqrt{-1}$.

If the two hop sizes $R_a$ and $R_s$ differ, i.e. if $R_a \neq R_s$, then the phases of $Y(m, k)$ need to be modified such that “horizontal phase coherence” is achieved. This means that for a sinusoid of constant frequency, successive frames overlap coherently without discontinuities or phase cancellation (destructive interference).

The phase vocoder approach is appropriate for polyphonic inputs, e.g. musical recordings. Its drawback is that the modification of the phase can produce an artefact known as "transient smearing", i.e. the temporal envelope of the signal is modified such that note onsets are perceived as having less attack and sound less percussive. Additional processing can be applied to the output phase to mitigate the transient smearing, e.g. by applying a method called "phase locking" [2] or by resetting the phase during periods of silence [3].

Following a suitable procedure of modifying the phase, an output is obtained. The spectral coefficients of the output can be written with polar coordinates as $Y = |Y| \exp(j\Phi_y)$ where $\Phi_y$ denotes the modified phase. The process of computing the phase $\Phi_y$ is in the following referred to as phase adaptation (PA).

Various ways to process two-channel audio input signals are known:

One option is to downmix the multi-channel signal to a single-channel signal, i.e. adding scaled versions of all channels, and to process the single-channel (mono) signal. Processing a mono mixdown of the input signal has the disadvantage that the stereophonic information is lost and thereby the sound quality is reduced.

A different option is to process the separated input channel signals independently. The main disadvantage of separately processing each channel signal is that arbitrary decorrelation between the channels is introduced which distorts the stereo image. Because the phase adaptation of time scale modification is a signal dependent processing, the relations between the phases of the individual channels are not preserved when the corresponding channel signals are different. The distortions of the spatial information can be perceived as blurring or widening of the stereo image of direct sound sources (singers or soloist, for example).

An object of the invention is to improve existing time scale modification methods with respect to the quality of the spatial characteristics – e.g. the stereo image – of the processed signal.

The object is achieved by an apparatus as well as by a method.

The object is achieved by an apparatus for processing a multichannel audio signal. The multichannel audio signal is comprising a plurality of – or at least two – channel signals. The apparatus is configured for performing a time scale modulation of the multichannel audio signal, i.e. the multichannel audio signal is slowed down or sped up without affecting its pitch. The apparatus comprises a phase adaptor and a separator. The phase adaptor is configured for providing at least one processed signal by modifying a phase of a signal based on a combination of the channel signals. The separator is configured for providing separated signals based on the at least one processed signal.

The invention improves processing signals and especially audio or speech signals. Especially addressed is the problem of processing stereophonic input signals having two or more channels.

An advantage of the invention is that the spatial characteristics of the input audio signal are preserved such that the perceived stereo image is not distorted. In particular, the positions of the sound sources and the diffuseness are not changed due to the TSM processing.

The invention thus addresses the problem that the spatial characteristics of the output signal are severely distorted in the state of art, which is most noticable when listening to sound sources panned to the center of the stereo image.

The apparatus performs the time scale modification of the multichannel audio signal by performing a phase adaptation. In the state of art, different procedures of modifying the phase (including phase locking and other means) are known.

One embodiment for modifying the phase comprises to modify the phase such that the phase propagation between adjacent frames with the synthesis hop size $R_s$ is identical to the phase propagation of the input signal for the analysis hop size $R_a$. This ensures that the horizontal phase coherence (i.e. the temporal evolution of the phases in each frequency bin) is maintained. This is achieved by computing the instantaneous frequency at time frame m given the input phase of the current and the preceding time frame and the analysis hop size $R_a$ and the STFT (Short Time Fourier Transform) parameters. The desired phase propagation is computed using the instantaneous frequency and the synthesis hop size $R_s$.

In an additional embodiment, the foregoing method is enhanced by a method comprising "phase locking". Phase locking aims to improve vertical phase coherence, i.e. to maintain the relation of the phase between adjacent frequency bins in each frame. This improves the sound quality as is for example noticeable when processing music signals containing transients or percussive notes.

The phase adaptor is configured to adapt the phase of at least one combination of the channel signals comprised by the multichannel audio signal. For this, the signals to be processed by the phase adaptor are given with polar coordinates by a magnitude and a phase.

The separator following the phase adaptor provides separated signals based on the processed signal, i.e. based on the signal with a modified phase. The separator reverses the combination of signals and extracts or generates separated signals.

Hence, the apparatus modifies the phase of at least one combination of channel signals and provides individual modified signals by separating the phase adapted combination of channel signals into separated signals.

The inventive phase adaptation is most suitable as an extension to time scale modification using the phase vocoder, i.e. processing in the frequency domain. However, it can also be applied as an extension to TSM in the time domain. To this end, the time-domain TSM is configured to process the at least one signal that is obtained by combining the channel signals comprised by the multichannel audio signal. The separator following the time-domain TSM provides separated signals based on the processed signal.

In an embodiment, the phase adaptor is configured for providing N processed signals by modifying phases of N signals based on combinations of N channel signals. The separator is configured for providing N separated signals based on the processed signals. N is a number of channel signals comprised by the multichannel audio signal and is a integer greater than or equal to two. In an embodiment, N is the number of all channel signals comprised by the multichannel audio signal. Hence, N channel signals are converted into N combined signals that are – after the phase adaptation – separated into N signals.

According to an embodiment, the N combinations of the channel signals are linear combinations of the channel signals. In an embodiment, the phase adaptor handles various combinations of channel signals, whereas the combinations are linear combinations, e.g. the sum or the difference of channel signals.

In an embodiment, N equals two (N = 2) and the two combinations of the two channel signals are a sum and a difference signal. In a different embodiment, N is greater than two.

In a further embodiment, the apparatus comprises a transformer. The transformer is configured for providing transformed signals by transforming signals from the time domain into the frequency domain. The transformer allows to transform signals from the time into the frequency domain and, subsequently, to process these signals in the frequency domain. The signals to be transformed are in one embodiment the channel signals comprised by the multichannel audio signal.

According to an embodiment, the transformer is configured for providing transformed signals by applying a Short Time Fourier Transformation. The Short Time Fourier Transform or alternatively Short Term Fourier Transform (STFT), is a Fourier-related transform of local sections of a signal as it changes over time. In practice, a longer signal is divided into shorter segments of equal length and the Fourier transform is computed separately on each segment.

In a further embodiment, the apparatus comprises a combiner, wherein the combiner is configured for providing combined signals based on the channel signals. The transformer is configured for providing transformed signals based on the combined signals by applying a transformation. Here, the channel signals are combined and the combinations are transformed into the frequency domain.

In a different embodiment, the sequence of combiner and transformer is reversed. In this embodiment, the combiner comprised by the apparatus is configured for providing combined signals based on transformed signals provided by the transformer. In this embodiment, the combiner receives transformed signals from the transformer and combines them in order to provide combined signals. Thus, the channel signals are individually transformed into the frequency domain and are combined in this domain by the combiner.

In a further embodiment, the combiner is configured for providing a sum signal as a combined signal by calculating a sum of two transformed signals. In this embodiment, a linear combination of the channel signals via their transformed signals in the form of a summation is used to provide at least one combined signal. If the multichannel audio signal comprises, for example, two channel signals as stereo signals, then both channel signals are added to each other for obtaining one combined signal being a sum signal. In a different embodiment, the combiner calculates a sum of channel signals of at least two channel signals.

According to an embodiment, the combiner is configured for providing a difference signal as combined signal by calculating a difference between two transformed signals. In this embodiment, the linear combination of signals is a difference between two signals. The signals to be combined are here transformed signals. In a different embodiment, the combiner calculates a difference between two channel signals.

According to an embodiment, the combiner is configured for providing combined signals by applying a mixing matrix – called **g** – having dimensions N times N to a multichannel signal comprising the N transformed signals, according to $\tilde{\mathbf{X}} = \mathbf{g}\mathbf{X}$.

N is a number of channel signals comprised by the multichannel audio signal and $\tilde{\mathbf{X}}$ is the matriced multichannel signal based on the transformed signals and represents, thus, the combined signals in a general form.

With other words: the combiner is configured for providing the combined signals by applying a mixing matrix having dimensions N times N to the N transformed signals – preferable in a matriced format – wherein the transformed signals are based on the N channel signals belonging to the multichannel audio signal.

As the mixing matrix has the same number of columns and rows, the number of signals to be combined equals the number of combined signals, as shown in the following equation:

$$\begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_N \end{bmatrix} = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,N} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,N} \\ \vdots & & & \\ g_{N,1} & g_{N,2} & \cdots & g_{N,N} \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

Applying the mixing matrix is equivalent to computing the k-th channel signal of the combined signal by summing all channel signals of the multichannel audio signal multiplied by the corresponding elements of the k-th row of the matrix, as shown in the following equation:

$$\tilde{X}_k = g_{k,1}X_1 + g_{k,2}X_2 + \ldots g_{k,N}X_N$$

In a different embodiment, the defined mixing matrix is applied to a signal based on the N channel signals. The mixing matrix allows to combine any given number of signals, being either transformed or channel signals.

In a further embodiment, the phase adaptor is configured for modifying the phases by applying a phase vocoder method. A phase vocoder is a vocoder scaling both the frequency and time domains of audio signals by using phase information.

According to an embodiment, the phase adaptor is configured for providing processed signals by their polar coordinates having amplitudes and modified phases. Thus, in one embodiment with N = 2 channel signals $Y_s$, $Y_d$, the processed signals are given by: $Y_{s/d} = |Y_{s/d}| \exp(j\Phi_{s/d})$ with $\Phi_{s/d} = \arg Y_{s/d}$ and $j = \sqrt{-1}$.

The separator provides single signals based on the processed signals that are based on the combined signals. The signals provided by the separator are to be used as time scale modified channel signals. Hence, according to an embodiment if the multichannel audio signals comprises N channel signals, then the separator provides N separated signals.

According to an embodiment, the separator is configured for providing a separated signal based on a difference between two of the processed signals.

In a further embodiment, the separator is configured for providing a separated signal based on a sum of two of the processed signals.

In an embodiment, the separator applies a factor – for example 0,5 – to the amplitude of the respective combination of the processed signals.

According to an embodiment, the separator is configured for providing N separated signals by applying an inverse mixing matrix having dimensions N times N to a matriced signal based on the N processed signals. N is a number of channel signals comprised by the multichannel audio signal. This embodiment allows to process any given number of processed signals and to provide the corresponding number of separated signals. The inverse mixing matrix is the inverse of the mixing matrix used for obtaining the combined signals.

According to an embodiment, the apparatus comprises a corrector, wherein the corrector is configured to modify the separated signals by replacing amplitudes of the separated signals with amplitudes based on amplitudes of the channel signals comprised by the multichannel audio signal.

This embodiment comprises to process a linear combination of the channel signals and to apply a magnitude correction procedure after the signals have been modified in the frequency domain.

The novelty of the presented method is, thus, in this embodiment two-fold:

1) to apply the TSM processing to the preferably linear combinations, e.g. to the sum signal and the difference signal and
2) to apply a processing for restoring the magnitude of the output signal in the time-frequency domain in order to restore the inter-channel level differences (ICLD).

In a further embodiment, the corrector is configured to replace the amplitudes of the separated signals with amplitudes of the corresponding transformed signals provided by a transformer, i.e. by the amplitudes of the channel signals in the frequency domain.

The corrector, thus, replaces the amplitudes of the separated signals having the adapted phase by the amplitudes of the corresponding signals before the adaptation. Hence, the original amplitudes are restored.

The following embodiments take care of signals in the frequency domain and allow to process them in the time domain.

In an embodiment, the apparatus comprises an inverse transformer. The inverse transformer is configured for providing modified channel signals based on the separated signals by applying an inverse transformation. This inverse transformation transforms in an embodiment signals from the frequency into the time domain.

According to another embodiment, the apparatus comprises an inverse transformer. The inverse transformer is configured for providing modified and corrected channel signals based on corrected signals provided by the corrector by applying an inverse transformation.

According to a different embodiment, the inverse transformer is configured for applying an inverse Short Time Fourier Transformation.

The inverse transformer is, thus, configured to reverse the kind of transformation performed in a step preceding the phase adaptation.

According to an embodiment, the apparatus comprises an extractor, wherein the extractor is configured for providing channel signals comprised by the multichannel audio signal. In this embodiment, the apparatus e.g. receives the multichannel audio signal and the extractor provides the individual channel signals. In a different embodiment, the channel signals are separately submitted to the apparatus.

The following embodiment allows to facilitate the computational steps and the requirements for the units.

According to an embodiment, the apparatus is configured to perform steps on combinations of channel signals based on a difference with less precision than on different combinations of the channel signals.

The object is also achieved by a method for processing a multichannel audio signal.

The method comprises at least the following steps:
providing at least one combined signal based on channel signals comprised by the multichannel audio signal,
providing a processed signal by performing a time scale modulation of the combined signal, and
providing modified channel signals based on a separation of the processed signals.

The time scale modulation is in one embodiment performed by a phase adaptation.

According to an embodiment, the method further comprises modifying amplitudes of modified channel signals by replacing the amplitudes with amplitudes based on amplitudes of the corresponding channel signals.

In this embodiment, the following steps happen: The channel signals are combined into combined signals. The combined signals or signals based on the combined signals undergo a phase adaptation in order to perform the time scale modification. The phase adapted signals are separated in separate signals. The signals comprise phases and amplitudes. The amplitudes/magnitudes of these signals are replaced by amplitudes based on the channel signals.

The embodiments of the apparatus can also be performed by steps of the method and corresponding embodiments of the method. Therefore, the explanations given for the embodiments of the apparatus also hold for the method.

The object is also achieved by a computer program for performing, when running on a computer or a processor, the method of any of the preceding embodiments.

The invention will be explained in the following with regard to the accompanying drawings and the embodiments depicted in the accompanying drawings, in which:

Fig. 1 shows a block diagram of a time scale modification method in the frequency domain according to the state of art,

Fig. 2 illustrates a block diagram of the inventive time scale modification method for input audio signals with two channels,

Fig. 3 provides schematically an embodiment of the apparatus,

Fig. 4 provides a different embodiment of the apparatus and

Fig. 5 illustrates a general version of the embodiment of Fig. 4.

A block diagram of an embodiment of the inventive method is depicted in Fig. 2 for the example of input signals having two channels (e.g. the left and right channel of stereo sound).

The input audio signal is $\mathbf{x} = [x_1\ x_2]^T$, where $x_1$ denotes the first channel signal and $x_2$ denotes the second channel signal. A short-time Fourier transform (STFT) representation is computed for $x_1$ and $x_2$, yielding $X_1$ and $X_2$, respectively, in the step STFT.

According to the invention, a sum signal $X_s$ and a difference signal $X_d$ are computed in the step named MS from the channel signals – here: the transformed channel signals $X_1$ and $X_2$ – of the multichannel audio signal according to

$$X_s = (X_1 + X_2), \tag{6}$$

$$X_d = (X_1 - X_2). \tag{7}$$

In a different embodiment, the order of STFT and the combination of the signals in step MS are switched. This takes into consideration, that computing the sum and difference signals can also be performed in the time domain and that the STFT can be computed from the time domain sum and difference signals. It is however advantageous to apply the order as shown in Fig. 2 to reduce the computational load.

The sum signal $X_s$ and the difference signal $X_d$ are then processed by means of a phase adaptation (PA) method, e.g. by using the phase vocoder method described in [2], or any other time scale modification method. The processed sum signal and difference signal are denoted by $Y_s$ and $Y_d$, respectively.

This is followed by an inverse transformation in step inverseMS, e.g. by transforming the signals $Y_s$ and $Y_d$ by an inverse short-time Fourier transform.

The processed signals for the first and the second channel are obtained in the shown embodiment by applying the following Equations (8) and (9), i.e. the inverse processing of Equations (6) and (7).

Hence, the two separated signals $Y_1$ and $Y_2$ are given in the shown embodiment by:

$$Y_1 = 0.5\,(Y_s + Y_d), \tag{8}$$

$$Y_2 = 0.5\,(Y_s - Y_d). \tag{9}$$

In a general form, the computation of the sum signal and the difference signal (as examples of the combined signals) are expressed in matrix notation as

$$\tilde{\mathbf{X}} = \mathbf{g}\mathbf{X} \tag{10}$$

with STFT coefficients of the input signal $\mathbf{X} = [X_1 \ldots X_N]^T$, a number of channels N, STFT coefficients of the matriced signal $\tilde{\mathbf{X}} = [\tilde{X}_1 \ldots \tilde{X}_N]^T$, and a mixing matrix g of size N times N.

For example, the sum signal $X_s$ and the difference signal $X_d$ – as given by Equations (6) and (7) – are obtained for N = 2 by setting $\mathrm{g} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\tilde{\mathbf{X}} = [X_s\ X_d]^T$.

The inverse processing of the matrix operation, i.e. computing the channel signals $\mathbf{Y} = [Y_1 \ldots Y_N]^T$ from the matriced signal $\tilde{\mathbf{Y}} = [\tilde{Y}_1 \ldots \tilde{Y}_N]^T$ is obtained from

$$\mathbf{Y} = \mathbf{g}^{-1}\tilde{\mathbf{Y}} \tag{11}$$

where $\tilde{\mathbf{Y}}$ is the result of the PA processing applied to $\tilde{\mathbf{X}}$ and $g^{-1}$ is the inverse of matrix g. With this generalization, the proposed method can also be applied to signals having more than two channels.

A very advantageous step of the shown embodiment is to correct the magnitude of the spectral coefficients such that the resulting complex-valued spectral coefficients have the phase of the result of the separation step using Equations (8) and (9) and the magnitude of $X_1$ and $X_2$ of the transformed channel signals.

The separated signals are given in polar coordinates as:

$Y_{1/2} = |Y_{1/2}| \exp(j\Phi_{y,1/2})$ with $\Phi_{y,1/2} = \arg Y_{1/2}$ and $j = \sqrt{-1}$.

Hence, for the shown two-channel input signals, the corrected signals $Z_1$ and $Z_2$ following the step MC are given by:

$$Z_1 = |X_1| \exp(j\Phi_{y,1}), \qquad (12)$$

$$Z_2 = |X_2| \exp(j\Phi_{y,2}). \qquad (13)$$

This step ensures that the inter-channel level differences (ICLD) of the audio signal is preserved.

Hence, in this step the magnitudes of the processed signals are substituted by the original amplitudes of the transformed channel signals.

For the general case of input signals having more than two channels, each output signal after the inverse matrix operation – i.e. after the separation in separated signals following the individual phase adaptation – is modified in an embodiment such that its magnitude is replaced by the magnitude of the corresponding – preferably transformed – channel signal (before matricing, i.e. before computing the combinations of different channel signals).

The PA processing has a considerable computational complexity. It involves various processing steps that can be implemented with reduced precision in order to reduce the computational load.

For example, the computation of the polar coordinates (magnitude and phase) of a complex number given by its Cartesian coordinates (real and imaginary component) can be implemented with lower or higher precision. Often, computations at lower precision have lower computational costs but introduce an error.

The computational load can be reduced by taking advantage of the fact that for typical audio signals (e.g. musical recordings or broadcast signals) the sum signal has higher energy then the difference signal. Errors that result from approximations in the computation have a smaller effect when they are introduced in the difference signal and greater effect when they occur in the sum signal. This can be exploited by applying methods with smaller precision for computing the difference signal $Y_d$ and methods with larger precision when computing the sum signal $Y_s$.

Another means for reducing the computational load is to skip certain processing steps. For example, the phase locking can be skipped when computing the difference signal $Y_d$. Phase locking refers to an additional processing step for improving the sound quality. For reducing the computational load, the phases for the difference signal $Y_d$ are computed such that the phase propagation is maintained without applying the phase locking processing.

Another means for reducing the computational load is to apply the processing with high-quality (large precision including all processing steps like phase locking) only up to a maximum frequency value. For a digital signal sampled at 48 kHz, for example, the high quality processing is only applied to frequency bands up to a maximum value of 10 kHz. The maximum frequency up to which the processing is applied with the best possible quality can be further reduced for the computation of the difference signal.

Fig. 3 shows an embodiment of the apparatus 1 which performs a time-scale modification of a multichannel audio signal 100. The multichannel audio signal 100 may comprise more than two channel signals. The time-scale modification is achieved by a phase adaption.

An extractor 2 retrieves the channel signals $x_1$, $x_2$ comprised by the multichannel audio signal 100. In the shown embodiment, there are just two channel signals. Nevertheless, the invention is not limited to two channel signals.

In a different – and not shown – embodiment, the channel signals $x_1$, $x_2$ are separately provided to the apparatus 1. Hence, for such an embodiment no extractor is required.

The channel signals $x_1$, $x_2$ are signals in the time domain and are, in the shown embodiment, submitted to a transformer 3. The transformer 3 transforms the channel signals $x_1$, $x_2$ into the frequency domain and, thus, provides transformed signals $X_1$, $X_2$.

The transformed signals $X_1$, $X_2$ are submitted to a combiner 4. The combiner 4 combines the transformed signals $X_1$, $X_2$ – in the given example – by applying linear combinations of them, e.g. by calculating a sum $X_s$ of the transformed signals and a difference $X_d$ between them. The number of combined signals is in one embodiment identical to the number of channel signals.

In a different – not shown – embodiment, the sequence of the transformer 3 and the combiner 4 is reversed. This implies that the combiner 4 combines the channel signals and the transformer 3 transforms the combined signals in this different and not shown embodiment.

In the depicted embodiment, the combined signals – being combined transformed channel signals – $X_s$ and $X_d$ are submitted to the phase adaptor 5.

The phase adaptor 5 modifies the phases of the combined signals $X_s$ and $X_d$ and provides processed signals $Y_s$ and $Y_d$. The processed signals $Y_s$ and $Y_d$ have adapted phases reflecting the appropriate time scale modification of the combined signals $X_s$ and $X_d$. Hence, the combined signals are either slowed down or sped up.

In order to obtain time scale modified channel signals, the processed signals $Y_s$ and $Y_d$ are separated by the separator 6 providing separated signals $Y_1$ and $Y_2$ that undergo an inverse transformation by an inverse transformer 7. The resulting modified channel signals $y_1$ and $y_2$ are time signals and have the desired time scale.

Fig. 4 shows a different embodiment of the apparatus 1.

The structure of the apparatus 1 shown in Fig. 4 is similar to the embodiment shown in Fig. 3. The difference between both embodiments is given by the units following the separator 6. For the explanation of the other elements see the description of Fig. 3.

The separator 6 in the embodiment shown in Fig. 4 also provides the separated signals $Y_1$ and $Y_2$. These frequency domain signals $Y_1$ and $Y_2$ are submitted in this embodiment to a corrector 8 that is upstream to the inverse transformer 7.

The corrector 8 replaces the amplitudes of the separated signals $Y_1$ and $Y_2$ by the amplitudes of the corresponding transformed signals $X_1$ and $X_2$, i.e. with the amplitudes or magnitudes before the phase adaptation and especially before the combination of the channel signals.

The resulting corrected or amplitude modified signals $Z_1$ and $Z_2$ (compare equations (12) and (13)) are submitted to the inverse transformer 7 and are transformed into the time domain as modified and corrected channel signals $z_1$, $z_2$.

In order to enable the correction, the transformer 3 is connected with the corrector 8.

In a further, not shown embodiment, the sequence of the transformer 3 and the combiner 4 are switched and the transformer 3, thus, transforms the combined signals. For the correction of the separated signals $Y_1$, $Y_2$ the corrector 8 refers therefore to additional transformations of the corresponding channel signals $x_1$, $x_2$.

The embodiment of Fig. 5 is a generalized version of the one shown in Fig. 4.

Here, the multichannel audio signal 100 comprises N channel signals $x_1$, $x_2$, ... $x_N$ where N is a integer greater than two.

The channel signals $x_1$, $x_2$, ... $x_N$ being time signals retrieved by the extractor 2 are submitted to the transformer 3 providing the transformed signals in the frequency domain $X_1$, $X_2$,... $X_N$ which are here given by a vector $\vec{X}$ (alternatively given by **X**).

The following combiner 4 provides a vector of linear combinations $\tilde{\mathbf{X}} = [\tilde{X}_1 \; \ldots \tilde{X}_N]^T$ by applying equation (10).

It follows the phase adaptor 5 providing processed signals (here given by a vector: $\tilde{Y}$) that are separated by the separator 6. The separated signals $Y_1$, $Y_2$, ... $Y_N$ (here given by a vector $\vec{Y}$) are corrected with regard to their amplitudes by the corrector 8. The corrected signals $Z_1$, $Z_2$ , ... $Z_N$ (given here by a vector $\vec{Z}$) are submitted to the inverse transformer 7 yielding N modified and corrected channel signals $z_1$, $z_2$, ... $z_N$. The switch from upper case to lower case indicated the transformation from the frequency to the time domain.

Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a programmable computer or an electronic circuit. In some embodiments, some one or more of the most important method steps may be executed by such an apparatus.

The inventive transmitted or encoded signal can be stored on a digital storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disc, a DVD, a Blu-Ray, a CD, a ROM, a PROM, and EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may, for example, be stored on a machine readable carrier.

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier.

In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

A further embodiment of the inventive method is, therefore, a data carrier (or a non-transitory storage medium such as a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein. The data carrier, the digital storage medium or the recorded medium are typically tangible and/or non-transitory.

A further embodiment of the invention method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may, for example, be configured to be transferred via a data communication connection, for example, via the internet.

A further embodiment comprises a processing means, for example, a computer or a programmable logic device, configured to, or adapted to, perform one of the methods described herein.

A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

A further embodiment according to the invention comprises an apparatus or a system configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example, be a computer, a mobile device, a memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver.

In some embodiments, a programmable logic device (for example, a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

References

[1] M. Dolson, “The Phase Vocoder: A Tutorial”, Computer Music Journal, vol. 10, pp. 14 – 27, 1986.

[2] J. Laroche and M. Dolson, “Improved Phase Vocoder Time-Scale Modification of Audio”, IEEE Transaction on Speech and Audio Processing, vol. 7, no. 3, pp. 323 – 332, 1999.

[3] T. Karrer, E. Lee, and J. Borchers, “PhaVoRIT: A Phase Vocoder for Real-Time Interactive Time-Stretching,” in Proc. of ICMC, 2006.

Claims

1. Apparatus (1) for processing a multichannel audio signal (100) comprising a plurality of channel signals ($x_1$, $x_2$), wherein the apparatus (1) is configured for performing a time scale modulation of the multichannel audio signal (100), and wherein the apparatus (1) is comprising:

a phase adaptor (5), wherein the phase adaptor (5) is configured for providing at least one processed signal ($Y_s$, $Y_d$) by modifying a phase of a signal ($X_s$, $X_d$) based on a combination of the channel signals ($x_1$, $x_2$), and

a separator (6), wherein the separator (6) is configured for providing separated signals ($Y_1$, $Y_2$) based on the at least one processed signal ($Y_s$, $Y_d$).

2. Apparatus (1) of claim 1, wherein the phase adaptor (5) is configured for providing N processed signals ($Y_s$, $Y_d$) by modifying phases of N signals ($X_s$, $X_d$) based on combinations of N channel signals ($x_1$, $x_2$), wherein the separator (6) is configured for providing N separated signals ($Y_1$, $Y_2$) based on the processed signal ($Y_s$, $Y_d$), and wherein N is a number of channel signals ($x_1$, $x_2$) comprised by the multichannel audio signal (100).

3. Apparatus (1) of claim 2, wherein the N combinations of the channel signals ($x_1$, $x_2$) are linear combinations of the channel signals ($x_1$, $x_2$).

4. Apparatus (1) of any of claims 1 to 3, wherein the apparatus (1) comprises a transformer (3), and wherein the transformer (3) is configured for providing transformed signals ($X_1$, $X_2$) by transforming signals from the time domain into the frequency domain.

5. Apparatus (1) of claim 4,

wherein the transformer (3) is configured for applying a Short Time Fourier Transformation.

6. Apparatus (1) of any of claims 1 to 5, wherein the apparatus (1) comprises a combiner (4), wherein the combiner (4) is configured for providing combined signals ($X_s$, $X_d$) based on the channel signals ($x_1$, $x_2$), and wherein the transformer (3) is configured for providing transformed signals based on the combined signals by applying a transformation.

7. Apparatus (1) of claim 4 or 5, wherein the apparatus (1) comprises a combiner (4), and wherein the combiner (4) is configured for providing combined signals ($X_s$, $X_d$) based on transformed signals ($X_1$, $X_2$) provided by the transformer (3).

8. Apparatus (1) of claim 7, wherein the combiner (4) is configured for providing a sum signal ($X_s$) by calculating a sum of two transformed signals ($X_1$, $X_2$).

9. Apparatus (1) of claim 7 or 8, wherein the combiner (4) is configured for providing a difference signal ($X_d$) by calculating a difference between two transformed signals ($X_1$, $X_2$).

10. Apparatus (1) of claim 7, wherein the combiner (4) is configured for providing the combined signals ($X_s$, $X_d$) by applying a mixing matrix (g) having dimensions N times N to the N transformed signals ($X_1$, $X_2$) based on the N channel signals ($x_1$, $x_2$) belonging to the multichannel audio signal (100), and wherein N is a number of channel signals ($x_1$, $x_2$) comprised by the multichannel audio signal (100).

11. Apparatus (1) of any of claims 1 to 10, wherein the phase adaptor (5) is configured for modifying the phases by applying a phase vocoder method.

12. Apparatus (1) of any of claims 2 to 11,

wherein the separator (6) is configured for providing a separated signal ($Y_1$, $Y_2$) based on a difference between two of the processed signals ($Y_s$, $Y_d$).

13. Apparatus (1) of any of claims 2 to 12, wherein the separator (6) is configured for providing a separated signal ($Y_1$, $Y_2$) based on a sum of two of the processed signals ($Y_s$, $Y_d$).

14. Apparatus (1) of any of claims 2 to 13, wherein the separator (6) is configured for providing N separated signals ($Y_1$, $Y_2$) by applying an inverse mixing matrix ($g^{-1}$) having dimensions N times N to a matriced signal based on the N processed signals ($Y_s$, $Y_d$), and wherein N is a number of channel signals ($x_1$, $x_2$) comprised by the multichannel audio signal (100).

15. Apparatus (1) of any of claims 1 to 14, wherein the apparatus (1) comprises a corrector (8), and wherein the corrector (8) is configured to modify the separated signals ($Y_1$, $Y_2$) by replacing amplitudes of the separated signals ($Y_1$, $Y_2$) with amplitudes based on amplitudes of the channel signals ($x_1$, $x_2$).

16. Apparatus (1) of claim 16, wherein the corrector (8) is configured to replace the amplitudes of the separated signals ($Y_1$, $Y_2$) with amplitudes of transformed signals ($X_1$, $X_2$) provided by a transformer (3).

17. Apparatus (1) of any of claims 1 to 16, wherein the apparatus (1) comprises an inverse transformer (7), and wherein the inverse transformer (7) is configured for providing modified channel signals ($y_1$, $y_2$) based on the separated signals ($Y_1$, $Y_2$) by applying an inverse transformation.

18. Apparatus (1) of claim 15 or 16, wherein the apparatus (1) comprises an inverse transformer (7), and wherein the inverse transformer (7) is configured for providing modified and corrected channel signals ($z_1$, $z_2$) based on corrected signals ($Z_1$, $Z_2$) provided by the corrector (8) by applying an inverse transformation.

19. Apparatus (1) of claim 17 or 18, wherein the inverse transformer (7) is configured for applying an inverse Short Time Fourier Transformation.

20. Apparatus (1) of any of claims 1 to 19, wherein the apparatus (1) comprises an extractor (2), and wherein the extractor (2) is configured for providing channel signals ($x_1$, $x_2$) comprised by the multichannel audio signal (100).

21. Apparatus (1) of any of claims 1 to 20, wherein the apparatus (1) is configured to perform steps on combinations of channel signals ($x_1$, $x_2$) based on a difference between signals with less precision than on different combinations of signals.

22. Method for processing a multichannel audio signal (100), comprising: providing at least one combined signal based on channel signals comprised by the multichannel audio signal, providing a processed signal by performing a time scale modulation of the combined signal, and providing modified channel signals based on a separation of the processed signal.

23. Method of claim 22, further comprising: modifying amplitudes of modified channel signals by replacing the amplitudes with amplitudes based on amplitudes of corresponding channel signals.

24. Computer program for performing, when running on a computer or a processor, the method of claim 22 or 23.

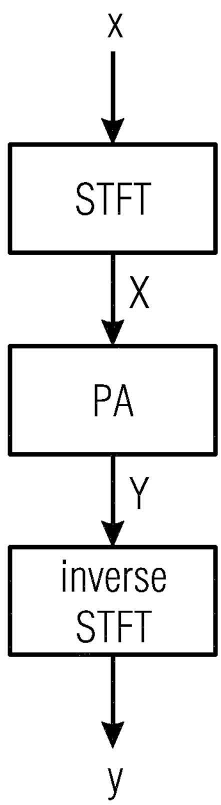


Fig. 1

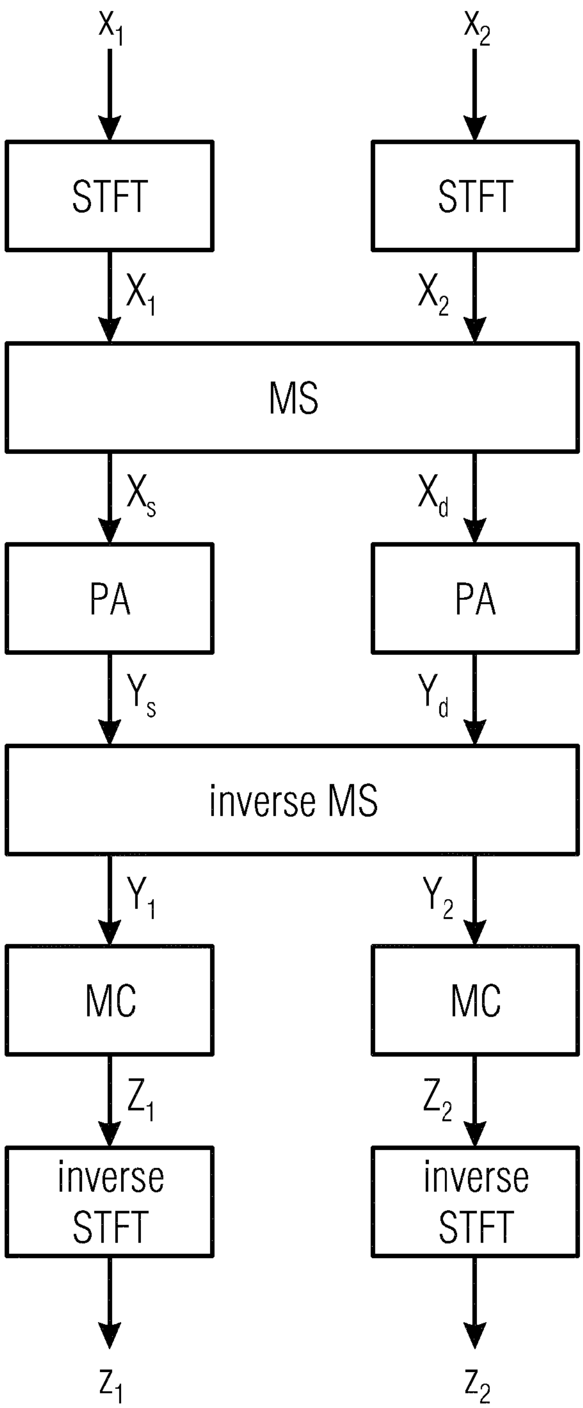
$x_1$
$x_2$
STFT
STFT
$X_1$
$X_2$
MS
$X_s$
$X_d$
PA
PA
$Y_s$
$Y_d$
inverse MS
$Y_1$
$Y_2$
MC
MC
$Z_1$
$Z_2$
inverse STFT
inverse STFT
$z_1$
$z_2$

Fig. 2

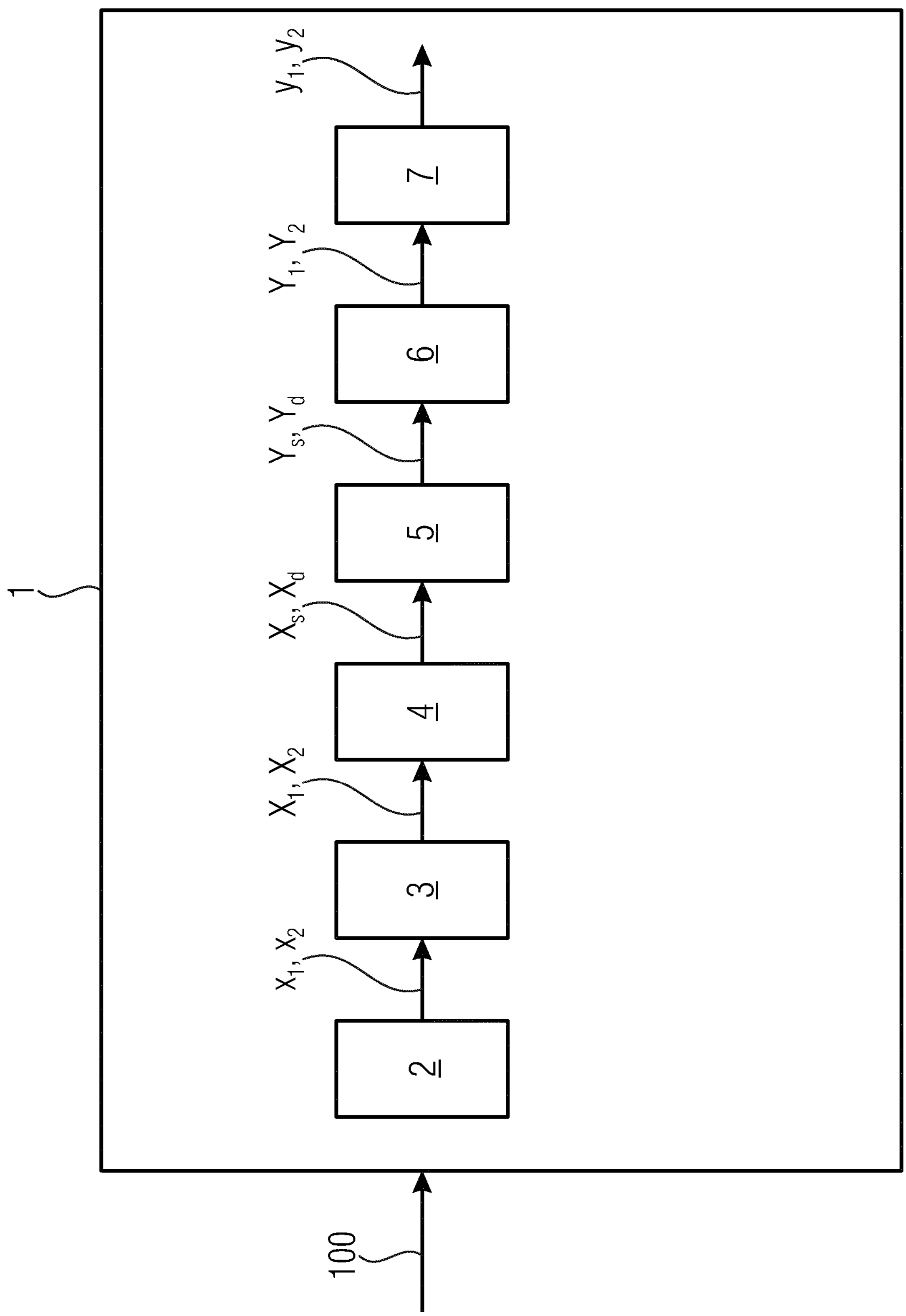


Fig. 3

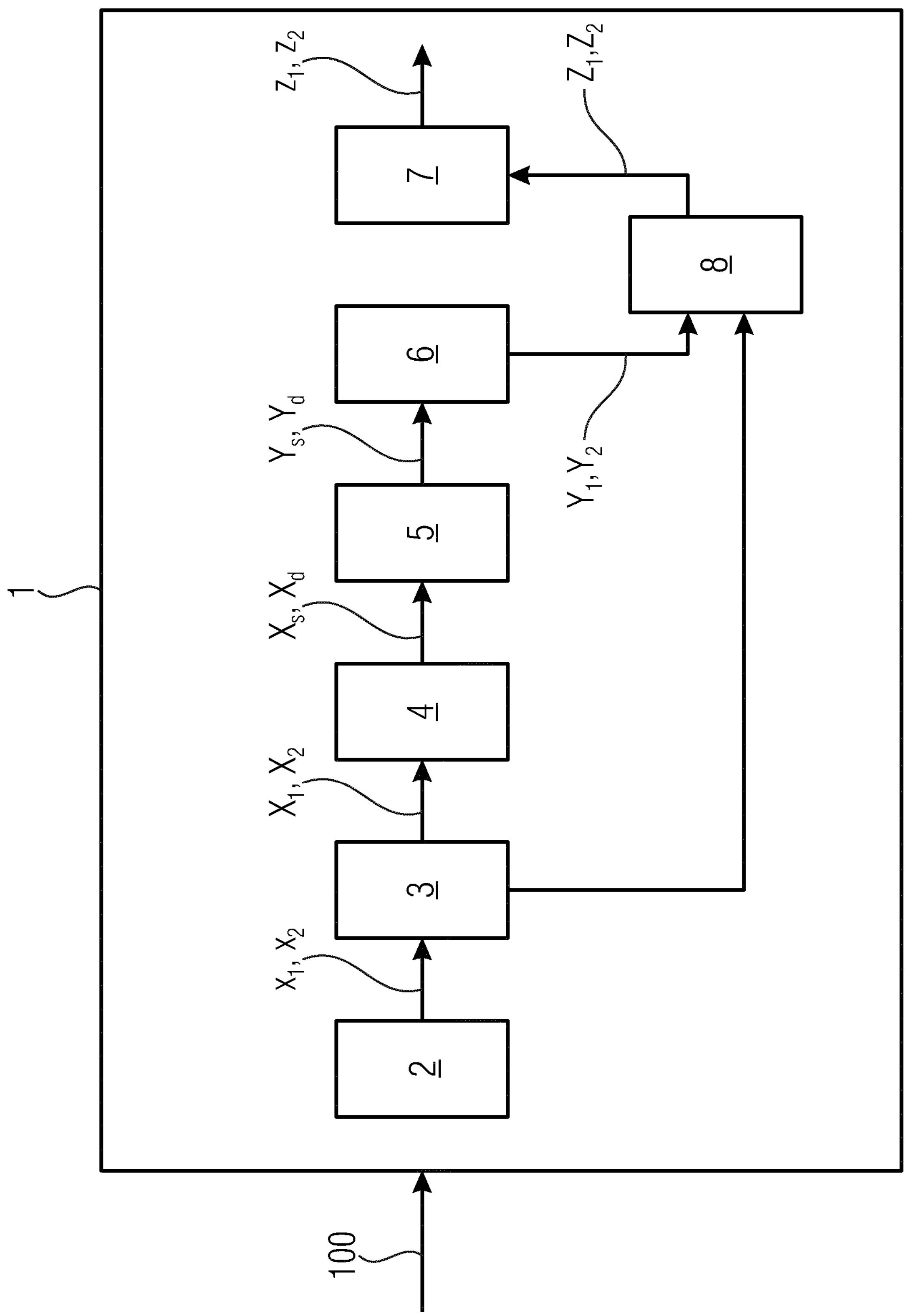


Fig. 4

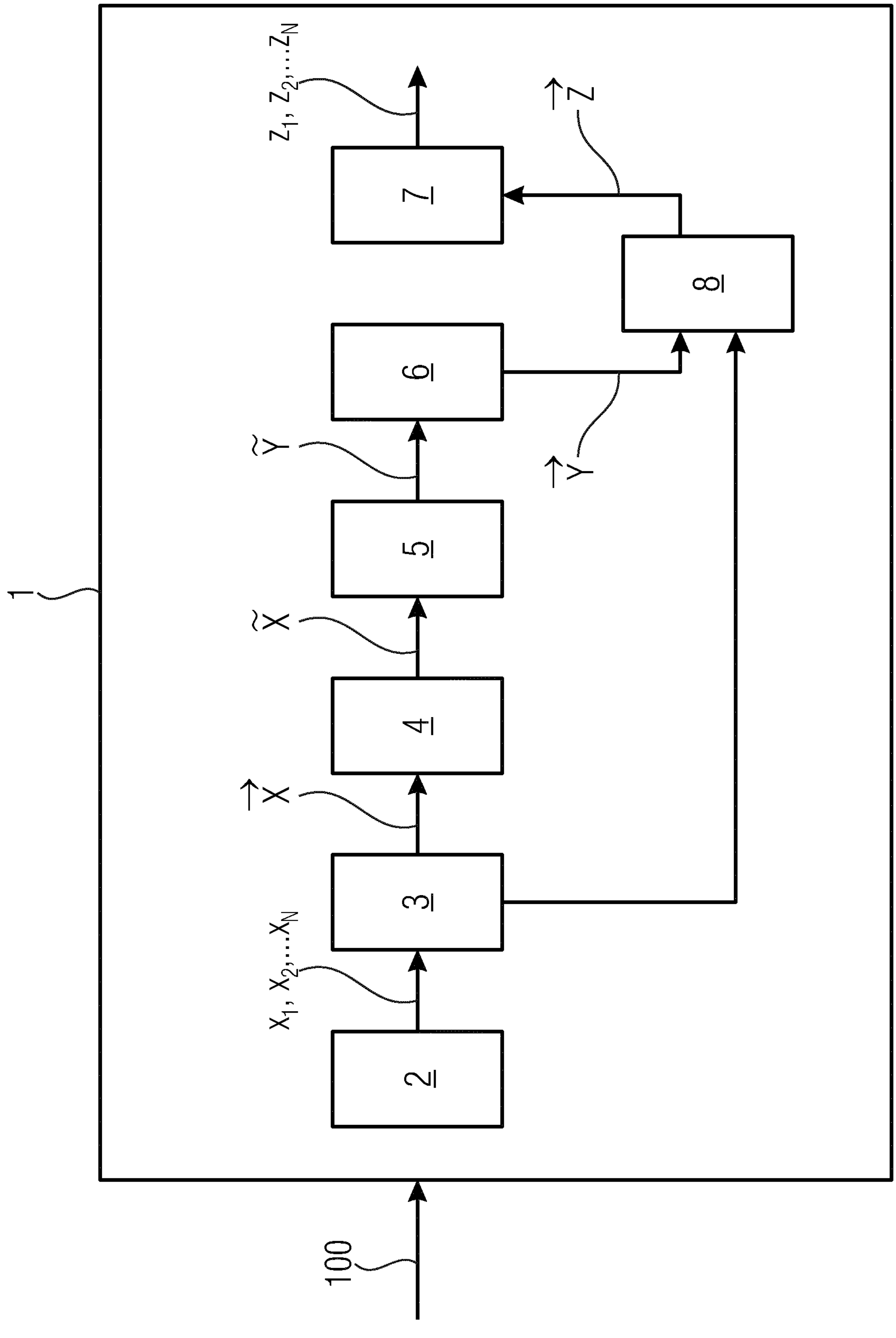
1
2
3
4
5
6
7
8
100
$x_1, x_2, \ldots x_N$
$\vec{X}$
$\tilde{X}$
$\tilde{Y}$
$\vec{Y}$
$\vec{Z}$
$Z_1, Z_2, \ldots Z_N$

Fig. 5

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/061895

**A. CLASSIFICATION OF SUBJECT MATTER**
INV. G11B27/28 G10L19/02 G11B27/00 G10L21/04
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
G11B G10L H04S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 2008/046967 A1 (NOKIA CORP [FI]; OJALA PASI [FI]; LAKANIEMI ARI [FI]; VIROLAINEN JUSSI) 24 April 2008 (2008-04-24)<br>page 9, lines 1-19; figures 1,4,5,6<br>page 8, lines 4-14<br>page 6, lines 14-36<br>page 8, lines 16-36<br>page 12, line 32 - page 13, line 3<br>----- | 1-24 |
| A | DOLSON M: "THE PHASE VOCODER: A TUTORIAL",<br>COMPUTER MUSIC JOUR, CAMBRIDGE, MA, US,<br>vol. 10, no. 4,<br>21 December 1986 (1986-12-21), pages 14-27, XP009029676,<br>ISSN: 0148-9267<br>cited in the application<br>the whole document<br>----- | 1-24 |

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 13 September 2017 | 22/09/2017 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Bruma, Cezar |

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | JEAN LAROCHE ET AL: "Improved Phase Vocoder Time-Scale Modification of Audio", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 7, no. 3, 1 May 1999 (1999-05-01), pages 323-332, XP011054370, ISSN: 1063-6676 cited in the application the whole document ----- | 1-24 |
| A | LEE E ET AL: "PhaVoRIT: A Phase Vocoder for Real-Time Interactive Time-Stretching", ICMC. INTERNATIONAL COMPUTER MUSIC CONFERENCE. PROCEEDINGS, XX, XX, 6 November 2006 (2006-11-06), pages 1-8, XP002424910, cited in the application the whole document ----- | 1-24 |
| A | US 2007/094031 A1 (CHEN JUIN-HWEY [US]) 26 April 2007 (2007-04-26) paragraphs [0125], [0126] ----- | 1-24 |
| A | WO 2012/167479 A1 (HUAWEI TECH CO LTD [CN]; TALEB ANISSE [DE]; VIRETTE DAVID [DE]; LI YUN) 13 December 2012 (2012-12-13) the whole document ----- | 1-24 |
| A | US 2005/137729 A1 (SAKURAI ATSUHIRO [JP] ET AL) 23 June 2005 (2005-06-23) the whole document ----- | 1-24 |
| A | FLANAGAN J L ET AL: "Phase vocoder", BELL SYSTEM TECHNICAL JOURNAL, AT AND T, SHORT HILLS, NY, US, vol. 45, no. 9, 1 November 1966 (1966-11-01), pages 1493-1509, XP011629282, ISSN: 0005-8580, DOI: 10.1002/J.1538-7305.1966.TB01706.X [retrieved on 2014-03-15] the whole document ----- -/-- | 1-24 |

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | LAROCHE J ET AL: "Phase-vocoder: about this phasiness business", APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 1997. 1997 I EEE ASSP WORKSHOP ON NEW PALTZ, NY, USA 19-22 OCT. 1997, NEW YORK, NY, USA,IEEE, US, 19 October 1997 (1997-10-19), page 4pp, XP010248209, ISBN: 978-0-7803-3908-8 the whole document ----- | 1-24 |
| A | PUCKETTE M: "Phase-locked vocoder", APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 1995., IEEE ASSP WORKSHOP ON NEW PALTZ, NY, USA 15-18 OCT. 1, NEW YORK, NY, USA,IEEE, US, 15 October 1995 (1995-10-15), pages 222-225, XP010154670, DOI: 10.1109/ASPAA.1995.482995 ISBN: 978-0-7803-3064-1 the whole document ----- | 1-24 |
| A | M. PORTNOFF: "Time-scale modification of speech based on short-time Fourier analysis", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING., vol. 29, no. 3, 1 June 1981 (1981-06-01), pages 374-390, XP055405797, USA ISSN: 0096-3518, DOI: 10.1109/TASSP.1981.1163581 the whole document ----- | 1-24 |
| A | PORTNOFF M R: "TIME-FREQUENCY REPRESENTATION OF DIGITAL SIGNALS AND SYSTEMS BASED ON SHORT-TIME FOURIER ANALYSIS", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, IEEE INC. NEW YORK, USA, vol. ASSP-28, no. 1, 1 February 1980 (1980-02-01), pages 55-69, XP000984575, ISSN: 0096-3518, DOI: 10.1109/TASSP.1980.1163359 the whole document ----- | 1-24 |

2

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2017/061895

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2008046967 A1 | 24-04-2008 | US 2008114606 A1<br>WO 2008046967 A1 | 15-05-2008<br>24-04-2008 |
| US 2007094031 A1 | 26-04-2007 | NONE | |
| WO 2012167479 A1 | 13-12-2012 | CN 103155030 A<br>EP 2710592 A1<br>JP 5734517 B2<br>JP 2014518407 A<br>US 2014140516 A1<br>WO 2012167479 A1 | 12-06-2013<br>26-03-2014<br>17-06-2015<br>28-07-2014<br>22-05-2014<br>13-12-2012 |
| US 2005137729 A1 | 23-06-2005 | NONE | |