

(19) **DANMARK**

(10) **DK/EP 3214620 T3**



Patent- og
Varemærkestyrelsen

(12) Oversættelse af
europæisk patentskrift

-
- (51) Int.Cl.: **G 10 L 25/60 (2013.01)** **G 10 L 21/02 (2013.01)** **G 10 L 21/0232 (2013.01)**
H 04 R 25/00 (2006.01)
- (45) Oversættelsen bekendtgjort den: **2019-11-25**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2019-09-18**
- (86) Europæisk ansøgning nr.: **17158062.4**
- (86) Europæisk indleveringsdag: **2017-02-27**
- (87) Den europæiske ansøgnings publiceringsdag: **2017-09-06**
- (30) Prioritet: **2016-03-01 EP 16157993**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark**
- (72) Opfinder: **Jensen, Jesper, c/o Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark**
De Haan, Jan Mark, c/o Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark
ANDERSEN, Asger Heidemann, c/o Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark
- (74) Fuldmægtig i Danmark: **Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark**
- (54) Benævnelse: **MONAURAL FORSTYRENDE TALEFORSTÅELIGHEDSFORUDSIGELSESENHED, ET HØREAPPARAT OG ET BINAURALT HØREAPPARATSYSTEM**
- (56) Fremdragne publikationer:
JALAL TAGHIA ET AL: "On mutual information as a measure of speech intelligibility", 2012 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP 2012) : KYOTO, JAPAN, 25 - 30 MARCH 2012 ; [PROCEEDINGS], IEEE, PISCATAWAY, NJ, 25 March 2012 (2012-03-25), pages 65-68, XP032227063, DOI: 10.1109/ICASSP.2012.6287818 ISBN: 978-1-4673-0045-2
Jesper Jensen ET AL: "Prediction of Intelligibility of Noisy and Time-Frequency Weighted Speech based on Mutual Information Between Amplitude Envelopes", interspeech 2013, 25 August 2013 (2013-08-25), XP055296103, Retrieved from the Internet: URL:http://vbn.aau.dk/files/196896029/Jens en_2013_.pdf [retrieved on 2016-08-17]
TANG YAN ET AL: "Evaluating the predictions of objective intelligibility metrics for modified and synthetic speech", COMPUTER SPEECH AND LANGUAGE, vol. 35, 25 March 2012 (2012-03-25), pages 73-92, XP029279568, ISSN: 0885-2308, DOI: 10.1016/J.CSL.2015.06.002
FEI CHEN ED - MITSIS GEORGIOS ET AL: "The relative importance of temporal envelope information for intelligibility prediction: A study on cochlear-implant vocoded speech", MEDICAL ENGINEERING & PHYSICS, BUTTERWORTH-HEINEMANN, GB, vol. 33, no. 8, 7 April 2011 (2011-04-07), pages 1033-1038, XP028267797, ISSN: 1350-4533, DOI: 10.1016/J.MEDENGPY.2011.04.004 [retrieved on 2011-04-12]

DESCRIPTION

SUMMARY

[0001] The present disclosure deals with monaural, intrusive intelligibility *prediction* of noisy/processed speech signals comprising a target signal component based on simultaneous knowledge of a substantially noise-free ('clean') version of the target signal component. The present disclosure further relates to a hearing aid comprising a monaural, intrusive intelligibility predictor unit, and to a binaural hearing aid system comprising first and second hearing aids, each comprising a monaural, intrusive intelligibility predictor unit, wherein the binaural hearing aid system is configured to establish a wireless link allowing the exchange of monaural speech intelligibility predictors or information derived therefrom between the first and second hearing aids.

[0002] Taghia et al., "On Mutual Information as a Measure of Speech Intelligibility", Proc. ICASSP 2012, deals with speech intelligibility prediction of noisy and processed noisy speech e.g. in hearing instruments.

[0003] Jensen and Taal, "Prediction of Intelligibility of Noisy and Time-Frequency Weighted Speech based on Mutual Information Between Amplitude Envelopes", Proc. Interspeech 2013, deals with the problem of predicting the average intelligibility of noisy and potentially processed speech signals, as observed by a group of normal hearing listeners.

[0004] Tang Yan et al., "Evaluating the predictions of objective intelligibility metrics for modified and synthetic speech", COMPUTER SPEECH AND LANGUAGE, vol. 35, 1 January 2016, pages 73-92, evaluates how well seven objective intelligibility metrics predict listener responses in three large datasets of modified and synthetic speech.

[0005] Fei Chen, "The relative importance of temporal envelope information for intelligibility prediction: A study on cochlear-implant vocoded speech", MEDICAL ENGINEERING & PHYSICS, vol. 33, no. 8, 7 April 2011, pages 1033-1038, describes factors influencing the intelligibility of cochlear implant listeners.

[0006] The invention is defined in the appended claims. All following occurrences of the word "embodiment(s)", if referring to feature combinations different from those defined by the independent claims, refer to examples which were originally filed but which do not represent embodiments of the presently claimed invention; these examples are still shown for illustrative purposes only.

A monaural speech intelligibility predictor unit:

[0007] In an aspect of the present application provides a monaural speech intelligibility predictor unit adapted for receiving a target signal comprising speech in an essentially noise-free version s and in a noisy and/or processed version x , the monaural speech intelligibility predictor unit being configured to provide as an output a final monaural speech intelligibility predictor value d indicative of a listener's perception of said noisy and/or processed version x of the target signal, the monaural speech intelligibility predictor unit comprising

1. a) A first input unit for providing a time-frequency representation $s(k,m)$ of said noise-free version s of the target signal, k being a frequency bin index, $k=1,2, \dots, K$, and m being a time index;
2. b) A second input unit for providing a time-frequency representation $x(k,m)$ of said noisy and/or processed version x of the target signal, k being a frequency bin index, $k=1, 2, \dots, K$, and m being a time index;
3. c) A first envelope extraction unit for providing a time-frequency sub-band representation $s_j(m)$ of the noise-free version s of the target signal representing temporal envelopes, or functions thereof, of frequency sub-band signals $s_j(m)$ of said noise-free target signal, j being a frequency sub-band index, $j=1, 2, \dots, J$, and m being the time index;
4. d) A second envelope extraction unit for providing a time-frequency sub-band representation $x_j(m)$ of the noisy and/or processed version x of the target signal representing temporal envelopes, or functions thereof, of frequency sub-band signals $x_j(m)$ of said noisy and/or processed version of the target signal, $j=1, 2, \dots, J$, and m being the time index;
5. e) A first time-frequency segment division unit for dividing said time-frequency sub-band representation $s_j(m)$ of the noise-free version s of the target signal into time-frequency segments S_m corresponding to a number N of successive samples of said sub-band signals;
6. f) A second time-frequency segment division unit for dividing said time-frequency sub-band representation $x_j(m)$ of the noisy and/or processed version x of the target signal into time-frequency segments X_m corresponding to a number N of successive samples of said sub-band signals;
7. g) An intermediate speech intelligibility calculation unit adapted for providing intermediate speech intelligibility coefficients d_m estimating an intelligibility of said time-frequency segment X_m , said intermediate speech intelligibility coefficients d_m being based on said essentially noise-free, optionally normalized and/or transformed, time frequency segments $\tilde{S}_m$, and said noisy and/or processed, optionally normalized and/or transformed, time-frequency segments $\tilde{X}_m$;
8. h) A final monaural speech intelligibility calculation unit for calculating a final monaural speech intelligibility predictor d estimating an intelligibility of said noisy and/or processed version x of the target signal by combining said intermediate speech intelligibility coefficients d_m , or a transformed version thereof, over time.

[0008] In an embodiment, the monaural speech intelligibility predictor unit comprises a normalization and transformation unit adapted for providing normalized and/or transformed versions $\tilde{X}_m$ ($\tilde{S}_m$) of said time-frequency segments $X_m(S_m)$.

[0009] In an embodiment, the normalization and transformation unit is configured to apply one or more algorithms for row and/or column normalization and/or transformation operations to the time-frequency segments S_m and/or X_m . In an embodiment, the normalization and transformation unit is configured to provide at least one normalization and/or transformation operation of rows and at least one normalization and/or transformation operation of columns to the time-frequency segments S_m and/or X_m .

[0010] In an embodiment, the monaural speech intelligibility predictor unit comprises a normalization and transformation unit configured to provide normalization and/or transformation of rows and columns of the time-frequency segments S_m and X_m , wherein the normalization and/or transformation of rows comprise(s) at least one of the following operations

R1) mean normalization of rows (cf. row normalization g_1 below),

R2) unit-norm normalization of rows (cf. row normalization g_2 below),

R3) Fourier transform of rows (cf. row transformation g_3 below),

R4) providing a Fourier magnitude spectrum of rows (cf. row transformation g_4 below), and

R5) providing the identity operation (cf. row transformation g_5 below),

and wherein said normalization and/or transformation of columns comprises at least one of the following operations

C1) mean normalization of columns (cf. column normalization h_1 below), and

C2) unit-norm normalization of columns (cf. column normalization h_2 below).

[0011] In an embodiment, the final monaural speech intelligibility calculation unit is configured to combine said intermediate speech intelligibility coefficients d_m , or a transformed version thereof, by averaging over time, or by applying a MIN or MAX-function, or other algebraic or statistical function, to the intermediate speech intelligibility coefficients d_m , or a transformed version thereof.

[0012] In an embodiment, the first and second input units are configured to receive the noise free version of the target signal s (also termed the 'clean (version of the) target signal') and the noisy and/or processed version x of the target signal (termed the 'information signal x '),

respectively, as a time variant (time domain/full band) signal $s(n)$ and $x(n)$, respectively, n being a time index. In an embodiment, the first and second input units are configured to receive the clean target signal s and the information signal x , respectively, in a time-frequency representation $s(k,m)$ and $x(k,m)$, respectively, from another unit or device, k and m being frequency and time indices, respectively. In an embodiment, the first and second input units each comprises a frequency decomposition unit for providing a time-frequency representation $s(k,m)$ and $x(k,m)$ of the clean target signal s and the information signal x from a time domain version of the respective signals ($s(n)$ and $x(n)$, n being a time index). In an embodiment, the frequency decomposition unit comprises a band-pass filterbank (e.g., a Gamma-tone filter bank), or is adapted to implement a Fourier transform algorithm (e.g. a short-time Fourier transform (STFT) algorithm).

[0013] In an embodiment, the monaural speech intelligibility predictor unit comprises a voice activity detector unit for indicating whether or not or to what extent a given time-segment of the essentially noise-free version s and the noisy and/or processed version x , respectively, of the target signal comprises or is estimated to comprise speech, and providing a voice activity control signal indicative thereof. In an embodiment, the voice activity detector unit is configured to provide a binary indication identifying segments comprising speech or no speech. In an embodiment, the voice activity detector unit is configured to identify segments comprising speech with a certain probability. In an embodiment, the voice activity detector is applied to a time-domain signal (or full-band signal, $s(n)$, $x(n)$, n being a time index). In an embodiment, the voice activity detector is applied to a time-frequency representation of a signal ($s(k,m)$, $x(k,m)$, or $s_j(m)$, $x_j(m)$, k and j being frequency indices (bin and sub-band, respectively), m being a time index) or a signal originating therefrom). In an embodiment, the voice activity detector unit is configured to identify time-frequency segments comprising speech on a time-frequency unit level (or e.g. in a frequency sub-band signal $x_j(m)$). In an embodiment, the monaural speech intelligibility predictor unit is adapted to receive (e.g. wirelessly receive) a voice activity control signal from another unit or device.

[0014] In an embodiment, the monaural speech intelligibility predictor unit comprises a voice activity detector unit for identifying time-segments of the essentially noise-free version s and the noisy and/or processed version x , respectively, of the target signal comprising or estimated to comprise speech, and wherein the monaural speech intelligibility predictor unit is configured to provide modified versions of the essentially noise-free version s and the noisy and/or processed version x , respectively, of the target signal comprising only such time segments comprising speech or being estimated to comprise speech.

[0015] In an embodiment, the first and second time-frequency segment division units are configured to base the generation of the time-frequency segments S_m and X_m , respectively, or normalized and/or transformed versions, $\tilde{S}_m$ and $\tilde{X}_m$, thereof on the voice activity control signal, e.g. to generate said time-frequency segments in dependence of the voice activity control signal, e.g. only if speech is indicated to be present, or if the probability that the time-frequency segment in question contains speech is larger than a predefined value, e.g. 0.5).

[0016] In an embodiment, the monaural speech intelligibility predictor unit comprises a hearing loss model unit configured to apply a frequency dependent modification of the said noisy and/or processed version x of the target signal reflecting a deviation from normal hearing, e.g. a hearing impairment, of a relevant ear of the listener to provide a modified noisy and/or processed version x of the target signal for use together with said essentially noise-free version s of the target signal as a basis for calculating the final monaural speech intelligibility predictor d .

[0017] In an embodiment, the hearing loss model unit is configured to add a statistically independent noise signal, which is spectrally shaped according to an audiogram of the relevant ear of the listener, to said noisy and/or processed version x of the target signal.

[0018] The first and second envelope extraction units are configured for extracting a temporal envelope $s_j(m)$ and $x_j(m)$ comprising J sub-bands ($j=1, 2, \dots, J$) of the clean target signal s and the information signal x , respectively, from said time-frequency representation $s(k,m)$ and $x(k,m)$ of the clean target signal s and the information signal x , respectively. In an embodiment, the first and second envelope extraction units each comprises an algorithm for implementing a Hilbert transform, or for low-pass filtering the magnitude of complex-valued STFT signals $s(k,m)$ and $x(k,m)$, etc.

[0019] In an embodiment, the monaural speech intelligibility predictor unit is adapted to extract said temporal envelope signals $x_j(m)$ and $s_j(m)$, respectively, as

$$z_j(m) = f\left(\sqrt{\sum_{k=k1(j)}^{k2(j)} |z(k,m)|^2}\right),$$

where z represent x or s , $j=1, \dots, J$ and $m=1, \dots, M$, $k1(j)$ and $k2(j)$ denote DFT bin indices corresponding to lower and higher cut-off frequencies of the j^{th} sub-band, J is the number of sub-bands, and M is the number of signal frames in the signal in question, and $f(\cdot)$ is a function.

[0020] In an embodiment, the function $f(\cdot)=f(w)$, where w represents

$$\left(\sqrt{\sum_{k=k1(j)}^{k2(j)} |z(k,m)|^2}\right),$$

is selected among the following functions

- $f(w)=w$ representing the identity
- $f(w)=w^2$ providing power envelopes,
- $f(w)=2 \cdot \log w$ or $f(w)=w^\beta$, $0 < \beta < 2$, allowing the modelling of the compressive non-linearity of the healthy cochlea,

or combinations thereof.

[0021] In an embodiment, the function $f(\cdot)=f(w)$, where w represents

$$\left(\sqrt{\sum_{k=k1(j)}^{k2(j)} |z(k,m)|^2}\right)$$

$$\left(\sqrt{\sum_{k \in k_l(j)} |z(k, m)|^2} \right)^{-1},$$

is selected among the following functions

- $f(w)=w^2$ providing power envelopes,
- $f(w)=2 \cdot \log w$ or $f(w)=w^\beta$, $0 < \beta < 2$, allowing the modelling of the compressive non-linearity of the healthy cochlea,

or combinations thereof.

[0022] In an embodiment, the first and second time-frequency segment division units are configured to divide said time-frequency representations $s_j(m)$ and $x_j(m)$, respectively, into segments in the form of spectrograms corresponding to N successive samples of all sub-band signals, wherein the m^{th} segment Z_m is defined by the $J \times N$ matrix

$$Z_m = \begin{bmatrix} z_1(m-N+1) & \cdots & z_1(m) \\ \vdots & & \vdots \\ z_J(m-N+1) & \cdots & z_J(m) \end{bmatrix},$$

where z (Z) represents s (S) or x (X).

[0023] In an embodiment, the monaural speech intelligibility predictor unit comprises

- A first optional normalization and/or transformation unit adapted for providing normalized and/or transformed versions $\tilde{S}_m$ of said time-frequency segments S_m ;
- A second optional normalization and/or transformation unit adapted for providing normalized and/or transformed versions $\tilde{X}_m$ of said time-frequency segments X_m ;

[0024] In an embodiment, the normalization and/or transformation unit is configured to apply one or more algorithms for row and/or column normalization and/or transformation to the time-frequency segments S_m and/or X_m , respectively.

[0025] In an embodiment, the normalization and/or transformation unit is configured to apply one or more of the following algorithms to the time-frequency segments X_m and S_m , respectively, commonly denoted Z_m , where sub-script, time index m is skipped for simplicity in the following expressions:

- R1) Normalization of rows to zero mean:

$$g_1(Z) = Z - \mu_z^r \mathbf{1}^T,$$

where

$$\mu_z^r$$

is a $J \times 1$ vector whose j^{th} entry is the mean of the j^{th} row of Z , hence the superscript r in

, r

μ_z ,

where $\mathbf{1}$ denotes an $N \times 1$ vector of ones, and where superscript T denotes matrix transposition;

- R2) Normalization of rows to unit-norm:

$$g_2(Z) = D^r(Z)Z,$$

where

$$D^r(Z) = \text{diag}\left(\left[1/\sqrt{Z(1,:)Z(1,:)^H} \quad \dots \quad 1/\sqrt{Z(J,:)Z(J,:)^H}\right]\right),$$

where $\text{diag}(\cdot)$ is a diagonal matrix with the elements of the arguments on the main diagonal, and where $Z(j,:)$ denotes the j 'th row of Z , such that $D^r(Z)$ is a $J \times J$ diagonal matrix with the inverse norm of each row on the main diagonal, and zeros elsewhere, the superscript H denotes Hermitian transposition, and pre-multiplication with $D^r(Z)$ normalizes the rows of the resulting matrix to unit-norm;

- R3) Fourier transformation applied to each row

$$g_3(Z) = ZF,$$

where F is an $N \times N$ Fourier matrix;

- R4) Fourier transformation applied to each row followed by computing the magnitude of the resulting complex-valued elements

$$g_4 = |ZF|$$

where $|\cdot|$ computes the element-wise magnitudes;

- R5) The identity operator

$$g_5(Z) = Z.$$

- C1) Normalization of columns to zero mean:

$$h_1(Z) = Z - \mathbf{1}\mu_z^c,$$

where

$$\mu_z^c$$

is a $N \times 1$ vector whose j 'th entry is the mean of the j 'th row of Z , and where $\mathbf{1}$ denotes a $J \times 1$ vector of ones;

- C2) Normalization of columns to unit-norm:

$$h_2(Z) = ZD^c(Z),$$

where

$$D^c(Z) = \text{diag}\left(\left[1/\sqrt{Z(:,1)^H Z(:,1)} \quad \dots \quad 1/\sqrt{Z(:,N)^H Z(:,N)}\right]\right),$$

where $Z(:, n)$ denotes the n 'th column of Z , such that $D^c(Z)$ is a diagonal $N \times N$ matrix with the inverse norm of each column on the main diagonal, and zeros elsewhere, and where a post-multiplication with $D^c(Z)$ normalizes the columns of the resulting matrix to unit-norm.

[0026] In an embodiment, the intermediate speech intelligibility calculation unit is adapted to determine the intermediate speech intelligibility coefficients d_m in dependence on a, e.g. linear, sample correlation coefficient $d(a,b)$ of the elements in two $K \times 1$ vectors a and b , $d(a,b)$ being

defined by:

$$d(a, b) = \frac{\sum_{k=1}^K (a(k) - \mu_a)(b(k) - \mu_b)}{\sqrt{\sum_{k=1}^K (a(k) - \mu_a)^2 (b(k) - \mu_b)^2}},$$

where

$$\mu_a = \frac{1}{K} \sum_{k=1}^K a(k)$$

and

$$\mu_b = \frac{1}{K} \sum_{k=1}^K b(k),$$

where k is the index of the vector entry and K is the vector dimension.

[0027] In an embodiment, a and b represent (e.g. any K) elements from time frequency segments S_m (or $\tilde{S}_m$) and X_m (or $\tilde{X}_m$), respectively.

[0028] In an embodiment, a and b represent elements from columns of time frequency segments S_m (or $\tilde{S}_m$) and X_m (or $\tilde{X}_m$), respectively. In an embodiment, a and b represent elements from rows of time frequency segments S_m (or $\tilde{S}_m$) and X_m (or $\tilde{X}_m$), respectively. In an embodiment, a and b represent all elements in time frequency segments S_m (or $\tilde{S}_m$) and X_m (or $\tilde{X}_m$), respectively.

[0029] In an embodiment, the intermediate intelligibility index d_m is defined as

1. 1) the average sample correlation coefficient of (some of) the (e.g. all) *columns* in S_m and X_m , or $\tilde{S}_m$ and $\tilde{X}_m$, respectively, i.e.,

$$d_m = \frac{1}{N} \sum_{n=1}^N d(\tilde{S}_m(:, n), \tilde{X}_m(:, n)),$$

n being a column index, or as

2. 2) the average sample correlation coefficient of (some of) the (e.g. all) *rows* in S_m and X_m , or $\tilde{S}_m$ and $\tilde{X}_m$, i.e.,

$$d_m = \frac{1}{J} \sum_{j=1}^J d(\tilde{S}_m(j, :)^T, \tilde{X}_m(j, :)^T),$$

j being a row index, or as

3. 3) the sample correlation coefficient of *all* elements in S_m and X_m , or $\tilde{S}_m$ and $\tilde{X}_m$, i.e.,

$$d_m = d(\tilde{S}_m(:), \tilde{X}_m(:)),$$

where the notation $S_m(:)$ and $X_m(:)$, or $\tilde{S}_m(:)$ and $\tilde{X}_m(:)$, represents $NJ \times 1$ vectors formed by stacking the columns of the respective matrices. In an embodiment, the number of columns N is 30 (or more). In an embodiment, the number of columns is 15 (or more). In an embodiment, $N \leq 30$ and $J \leq 15$.

[0030] In an embodiment, the final speech intelligibility calculation unit is adapted to calculate the final speech intelligibility predictor d from the intermediate speech intelligibility coefficients d_m , optionally transformed by a function $u(d_m)$, as an average over time of said noisy and/or processed version x of the target signal:

$$d = \frac{1}{M} \sum_{m=1}^M u(d_m)$$

where M represents the duration in time units of the speech active parts of said noisy and/or processed version x of the target signal. In an embodiment, the duration of the speech active parts of the noisy and/or processed version x of the target signal is defined as a (possibly accumulated) time period where the voice activity control signal indicates that the noisy and/or processed version x of the target signal comprises speech.

[0031] In an embodiment, the function $u(d_m)$ is defined as

$$u(d_m) = \log\left(\frac{1}{1-d_m^2}\right),$$

or as

$$u(d_m) = d_m.$$

A hearing aid:

[0032] In an aspect, a hearing aid adapted for being located at or in left and right ears of a user, or for being fully or partially implanted in the head of the user, the hearing aid comprising a monaural speech intelligibility predictor unit as described above, in the detailed description of embodiments, in the drawings and in the claims is furthermore provided by the present disclosure.

[0033] In an embodiment, the hearing aid is configured to adaptively modify the processing of an input signal to the hearing aid to maximize the final monaural speech intelligibility predictor d . to enhance the user's intelligibility of an output signal of the hearing aid presented to the user

[0034] In an embodiment, the hearing aid comprises

1. a) A number of input units IU_i , $i=1, \dots, M$, M being larger than or equal to one, each being configured to provide a time-variant electric input signal y_i representing a sound input received at an i^{th} input unit, the electric input signal y_i comprising a target signal component and a noise signal component, the target signal component originating from a target signal source;
2. b) A configurable signal processing unit for processing the electric input signals and providing a processed signal u , which is connected to the hearing loss model of the

- monaural speech intelligibility predictor unit;
3. c) An output unit for creating output stimuli configured to be perceivable by the user as sound based on an electric output either in the form of the processed signal u from the signal processing unit or a signal derived therefrom; and
 4. d) Antenna and transceiver unit for receiving a wireless signal s' comprising the target signal and for extracting an essentially noise-free version s of the target signal, which is connected to the monaural speech intelligibility predictor unit;
 5. e) Wherein the final speech intelligibility predictor d is fed to the configurable signal processing unit.

The hearing aid (e.g. the monaural speech intelligibility predictor unit) preferably comprises a hearing loss model configured to provide that the input signal to the monaural speech intelligibility predictor unit (e.g. the output of the configurable processing unit, cf. e.g. FIG. 5B) is modified to reflect a deviation of a user's hearing profile from a normal hearing profile, e.g. to reflect a hearing impairment of the user.

[0035] In an embodiment, the configurable signal processing unit is adapted to control or influence the processing of the respective electric input signals, or one or more signals originating therefrom (e.g. a resulting beamformed signal) based on said final speech intelligibility predictor d provided by the monaural speech intelligibility predictor unit.

[0036] In an embodiment, the configurable signal processing unit is adapted to control or influence the processing of the respective electric input signals based on said final speech intelligibility predictor d when the target signal component comprises speech, such as only when the target signal component comprises speech (as e.g. defined by a voice (speech) activity detector).

[0037] In an embodiment, the configurable signal processing unit is adapted to control or influence the processing of the respective electric input signals to maximize the final speech intelligibility predictor d .

[0038] In an embodiment, the hearing aid is adapted to provide a frequency dependent gain and/or a level dependent compression and/or a transposition (with or without frequency compression) of one or frequency ranges to one or more other frequency ranges, e.g. to compensate for a hearing impairment of a user.

[0039] In an embodiment, the output unit comprises a number of electrodes of a cochlear implant or a vibrator of a bone conducting hearing aid. In an embodiment, the output unit comprises an output transducer. In an embodiment, the output transducer comprises a receiver (loudspeaker) for providing the stimulus as an acoustic signal to the user. In an embodiment, the output transducer comprises a vibrator for providing the stimulus as mechanical vibration of a skull bone to the user (e.g. in a bone-attached or bone-anchored hearing aid).

[0040] In an embodiment, the input unit comprises an input transducer for converting an input sound to an electric input signal. In an embodiment, the input unit comprises a wireless receiver for receiving a wireless signal comprising sound and for providing an electric input signal representing said sound. In an embodiment, the hearing aid comprises a directional microphone system adapted to enhance a target acoustic source among a multitude of acoustic sources in the local environment of the user wearing the hearing aid. In an embodiment, the directional system is adapted to detect (such as adaptively detect) from which direction a particular part of the microphone signal originates.

[0041] In an embodiment, the hearing aid comprises an antenna and transceiver circuitry for wirelessly receiving a direct electric input signal from another device, e.g. a communication device or another hearing aid. In general, a wireless link established by antenna and transceiver circuitry of the hearing aid can be of any type. In an embodiment, the wireless link is used under power constraints, e.g. in that the hearing aid comprises a portable (typically battery driven) device.

[0042] In an embodiment, the hearing aid comprises a forward or signal path between an input transducer (microphone system and/or direct electric input (e.g. a wireless receiver)) and an output transducer. In an embodiment, the signal processing unit is located in the forward path. In an embodiment, the signal processing unit is adapted to provide a frequency dependent gain according to a user's particular needs. In an embodiment, the hearing aid comprises an analysis path comprising functional components for analyzing the input signal (e.g. determining a level, a modulation, a type of signal, an acoustic feedback estimate, etc.). In an embodiment, some or all signal processing of the analysis path and/or the signal path is conducted in the frequency domain. In an embodiment, some or all signal processing of the analysis path and/or the signal path is conducted in the time domain.

[0043] In an embodiment, the hearing aid comprises an analogue-to-digital (AD) converter to digitize an analogue input with a predefined sampling rate, e.g. 20 kHz. In an embodiment, the hearing aid comprises a digital-to-analogue (DA) converter to convert a digital signal to an analogue output signal, e.g. for being presented to a user via an output transducer.

[0044] In an embodiment, the hearing aid comprises a number of detectors configured to provide status signals relating to a current physical environment of the hearing aid (e.g. the current acoustic environment), and/or to a current state of the user wearing the hearing aid, and/or to a current state or mode of operation of the hearing aid. Alternatively or additionally, one or more detectors may form part of an *external* device in communication (e.g. wirelessly) with the hearing aid. An external device may e.g. comprise another hearing aid, a remote control, and audio delivery device, a telephone (e.g. a Smartphone), an external sensor, etc. In an embodiment, one or more of the number of detectors operate(s) on the full band signal (time domain). In an embodiment, one or more of the number of detectors operate(s) on band split signals ((time-) frequency domain).

[0045] In an embodiment, the hearing aid further comprises other relevant functionality for the

application in question, e.g. compression, noise reduction, feedback reduction, etc.

Use of a monaural speech intelligibility predictor unit:

[0046] In an aspect, use of a monaural speech intelligibility predictor unit as described above, in the detailed description of embodiments, in the drawings and in the claims in a hearing aid to modify signal processing in the hearing aid aiming at enhancing intelligibility of a speech signal presented to a user by the hearing aid is furthermore provided by the present disclosure. In an embodiment, use of a monaural speech intelligibility predictor unit in a hearing aid in a noisy environment is provided (e.g. a car telephony situation, or other listening situation where a (e.g. substantially clean version of the) target speech signal is received wirelessly and acoustic noise is present at the user's ears) to enhance a user's intelligibility of speech in a noisy environment. In an embodiment, use of a monaural speech intelligibility predictor unit in an active ear protection device is provided.

A method of providing a monaural speech intelligibility predictor:

[0047] In a further aspect, a method of providing a monaural speech intelligibility predictor for estimating a user's ability to understand an information signal x comprising a noisy and/or processed version of a target speech signal is provided. The method comprises

- providing a time-frequency representation $s(k,m)$ of said noise-free version s of the target signal, k being a frequency bin index, $k=1, 2, \dots, K$, and m being a time index;
- providing a time-frequency representation $x(k,m)$ of said noisy and/or processed version x of the target signal, k being a frequency bin index, $k=1, 2, \dots, K$, and m being a time index;
- providing a time-frequency sub-band representation $s_j(m)$ of the noise-free version s of the target signal representing temporal envelopes, or functions thereof, of frequency sub-band signals $s_j(m)$ of said noise-free target signal, j being a frequency sub-band index, $j=1, 2, \dots, J$, and m being the time index;
- providing a time-frequency sub-band representation $x_j(m)$ of the noisy and/or processed version x of the target signal representing temporal envelopes, or functions thereof, of frequency sub-band signals $x_j(m)$ of said noisy and/or processed version of the target signal, $j=1, 2, \dots, J$, and m being the time index;
- dividing said time-frequency sub-band representation $s_j(m)$ of the noise-free version s of the target signal into time-frequency segments S_m corresponding to a number N of successive samples of said sub-band signals;
- dividing said time-frequency sub-band representation $x_j(m)$ of the noisy and/or processed version x of the target signal into time-frequency segments X_m corresponding to a number N of successive samples of said sub-band signals;

- providing intermediate speech intelligibility coefficients d_m estimating an intelligibility of said time-frequency segment X_m , said intermediate speech intelligibility coefficients d_m being based on said essentially noise-free, normalized and/or transformed time frequency segments $\tilde{S}_m$, and said noisy and/or processed, normalized and/or transformed time-frequency segments $\tilde{X}_m$;
- calculating a final monaural speech intelligibility predictor d estimating an intelligibility of said noisy and/or processed version x of the target signal by combining, e.g. by averaging or applying a MIN or MAX-function, said intermediate speech intelligibility coefficients d_m , or a transformed version thereof, over time.

[0048] It is intended that some or all of the structural features of the device described above, in the 'detailed description of embodiments' or in the claims can be combined with embodiments of the method, when appropriately substituted by a corresponding process and vice versa. Embodiments of the method have the same advantages as the corresponding devices.

[0049] In an embodiment, the method comprises subjecting a speech signal (a signal comprising speech) to a hearing loss model configured to model imperfections of an impaired auditory system to thereby provide said information signal x . By subjecting the speech signal (e.g. signal x' in FIG. 3A) to a hearing loss model (HLM in FIG. 3A), the resulting information signal x can be used as an input to the speech intelligibility predictor (MSIP in FIG. 3A), thereby providing a measure of the intelligibility of the speech signal for an *unaided* hearing impaired person. In an embodiment, the hearing loss model is a generalized model reflecting a hearing impairment of an average hearing impaired user. In an embodiment, the hearing loss model is configurable to reflect a hearing impairment of a particular user, e.g. including a frequency dependent hearing loss (deviation of a hearing threshold from a(n average) hearing threshold of a normally hearing person). By subjecting a speech signal (e.g. signal y in FIG. 3C) to a signal processing (e.g. SPU in FIG. 3C) intended to compensate for the user's hearing impairment, AND to a hearing loss model (HLM in FIG. 3C) the resulting information signal x can be used as an input to the speech intelligibility predictor (cf. e.g. MSIP in FIG. 3C), thereby providing a measure d of the intelligibility of the speech signal for an *aided* hearing impaired person. Such scheme may e.g. be used to evaluate the influence of different processing algorithms (and/or modifications of processing algorithms) on the user's (estimated) intelligibility of the resulting information signal (cf. e.g. FIG. 3B) or be used to online optimization of signal processing in a hearing aid (cf. e.g. 3C).

[0050] In an embodiment, the method comprises adding noise to a target speech signal to provide said information signal x , which is used as input to the method of providing a monaural speech intelligibility predictor value. The addition of a predetermined (or varying) amount of noise to an information signal can be used to - in a simple way - emulate a hearing loss of a user (to provide the effect of a hearing loss model). In an embodiment, the target signal is modified (e.g. attenuated) according to the hearing loss of a user, e.g. an audiogram. In an embodiment, noise is added to a target signal AND the target signal is attenuated to reflect a

hearing loss of a user.

A binaural hearing (aid) system:

[0051] In an aspect, a (first) binaural hearing system comprising left and right hearing aids as described above, in the detailed description of embodiments and drawings and in the claims is furthermore provided.

[0052] In an embodiment, each of the left and right hearing aids comprises antenna and transceiver circuitry for allowing a communication link to be established and information to be exchanged between said left and right hearing aids.

[0053] In an embodiment, the binaural hearing system further comprises a binaural speech intelligibility prediction unit for providing a final binaural speech intelligibility measure $d_{binaural}$ of the predicted speech intelligibility of the user, when exposed to said sound input, based on the monaural speech intelligibility predictor values d_{left} , d_{right} of the respective left and right hearing aids.

[0054] In an embodiment, the final binaural speech intelligibility measure $d_{binaural}$ is determined as the maximum of the speech intelligibility predictor values d_{left} , d_{right} of the respective left and right hearing aids: $d_{binaural} = \max(d_{left}, d_{right})$. Thereby a relatively simple system is provided implementing a better ear approach. In an embodiment, the binaural hearing system is adapted to activate such approach when an asymmetric listening situation is detected or selected by the user, e.g. a situation where a speaker is located predominantly to one side of the user wearing the binaural hearing system, e.g. when sitting in a car.

[0055] In an embodiment, the respective configurable signal processing units of the left and right hearing aids are adapted to control or influence the processing of the respective electric input signals based on said final binaural speech intelligibility measure $d_{binaural}$. In an embodiment, the respective configurable signal processing units of the left and right hearing aids are adapted to control or influence the processing of the respective electric input signals to maximize said final binaural speech intelligibility measure $d_{binaural}$.

[0056] In an embodiment, the binaural hearing system further comprises an auxiliary device.

[0057] In an embodiment, the system is adapted to establish a communication link between the hearing aid(s) and the auxiliary device to provide that information (e.g. control and status signals, possibly audio signals) can be exchanged or forwarded from one to the other.

[0058] In an embodiment, the auxiliary device is or comprises a remote control for controlling functionality and operation of the hearing aid(s). In an embodiment, the function of a remote control is implemented in a SmartPhone, the SmartPhone possibly running an APP allowing to

control the functionality of the audio processing device via the SmartPhone (the hearing aid(s) comprising an appropriate wireless interface to the SmartPhone, e.g. based on Bluetooth or some other standardized or proprietary scheme).

An APP:

[0059] In a further aspect, a non-transitory application, termed an APP, is furthermore provided by the present disclosure. The APP comprises executable instructions configured to be executed on an auxiliary device to implement a user interface for a hearing aid or a hearing system described above in the 'detailed description of embodiments', and in the claims. In an embodiment, the APP is configured to run on cellular phone, e.g. a smartphone, or on another portable device allowing communication with said hearing aid or said hearing system.

A computer readable medium:

[0060] In an aspect, a tangible computer-readable medium storing a computer program comprising program code means for causing a data processing system to perform at least some (such as a majority or all) of the steps of any one of the methods described above, in the 'detailed description of embodiments' and in the claims, when said computer program is executed on the data processing system is furthermore provided by the present application.

[0061] By way of example, and not limitation, such computer-readable media can comprise RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium that can be used to carry or store desired program code in the form of instructions or data structures and that can be accessed by a computer. Disk and disc, as used herein, includes compact disc (CD), laser disc, optical disc, digital versatile disc (DVD), floppy disk and Blu-ray disc where disks usually reproduce data magnetically, while discs reproduce data optically with lasers. Combinations of the above should also be included within the scope of computer-readable media. In addition to being stored on a tangible medium, the computer program can also be transmitted via a transmission medium such as a wired or wireless link or a network, e.g. the Internet, and loaded into a data processing system for being executed at a location different from that of the tangible medium.

A data processing system:

[0062] In an aspect, a data processing system comprising a processor and program code means for causing the processor to perform at least some (such as a majority or all) of the steps of the any one of the methods described above, in the 'detailed description of embodiments' and in the claims is furthermore provided by the present application.

A computer program:

[0063] A computer program (product) comprising instructions which, when the program is executed by a computer, cause the computer to carry out (steps of) the method described above, in the 'detailed description of embodiments' and in the claims is furthermore provided by the present application.

Definitions:

[0064] In the present context, a 'hearing aid' refers to a device, such as e.g. a hearing instrument or an active ear-protection device or other audio processing device, which is adapted to improve, augment and/or protect the hearing capability of a user by receiving acoustic signals from the user's surroundings, generating corresponding audio signals, possibly modifying the audio signals and providing the possibly modified audio signals as audible signals to at least one of the user's ears. A 'hearing aid' further refers to a device such as an earphone or a headset adapted to receive audio signals electronically, possibly modifying the audio signals and providing the possibly modified audio signals as audible signals to at least one of the user's ears. Such audible signals may e.g. be provided in the form of acoustic signals radiated into the user's outer ears, acoustic signals transferred as mechanical vibrations to the user's inner ears through the bone structure of the user's head and/or through parts of the middle ear as well as electric signals transferred directly or indirectly to the cochlear nerve of the user.

[0065] The hearing aid may be configured to be worn in any known way, e.g. as a unit arranged behind the ear with a tube leading radiated acoustic signals into the ear canal or with a loudspeaker arranged close to or in the ear canal, as a unit entirely or partly arranged in the pinna and/or in the ear canal, as a unit attached to a fixture implanted into the skull bone, as an entirely or partly implanted unit, etc. The hearing aid may comprise a single unit or several units communicating electronically with each other.

[0066] More generally, a hearing aid comprises an input transducer for receiving an acoustic signal from a user's surroundings and providing a corresponding input audio signal and/or a receiver for electronically (i.e. wired or wirelessly) receiving an input audio signal, a (typically configurable) signal processing circuit for processing the input audio signal and an output means for providing an audible signal to the user in dependence on the processed audio signal. In some hearing aids, an amplifier may constitute the signal processing circuit. The signal processing circuit typically comprises one or more (integrated or separate) memory elements for executing programs and/or for storing parameters used (or potentially used) in the processing and/or for storing information relevant for the function of the hearing aid and/or for storing information (e.g. processed information, e.g. provided by the signal processing circuit), e.g. for use in connection with an interface to a user and/or an interface to a

programming device. In some hearing aids, the output means may comprise an output transducer, such as e.g. a loudspeaker for providing an air-borne acoustic signal or a vibrator for providing a structure-borne or liquid-borne acoustic signal. In some hearing aids, the output means may comprise one or more output electrodes for providing electric signals.

[0067] In some hearing aids, the vibrator may be adapted to provide a structure-borne acoustic signal transcutaneously or percutaneously to the skull bone. In some hearing aids, the vibrator may be implanted in the middle ear and/or in the inner ear. In some hearing aids, the vibrator may be adapted to provide a structure-borne acoustic signal to a middle-ear bone and/or to the cochlea. In some hearing aids, the vibrator may be adapted to provide a liquid-borne acoustic signal to the cochlear liquid, e.g. through the oval window. In some hearing aids, the output electrodes may be implanted in the cochlea or on the inside of the skull bone and may be adapted to provide the electric signals to the hair cells of the cochlea, to one or more hearing nerves, to the auditory cortex and/or to other parts of the cerebral cortex.

[0068] A 'hearing system' refers to a system comprising one or two hearing aids, and a 'binaural hearing system' refers to a system comprising two hearing aids and being adapted to cooperatively provide audible signals to both of the user's ears. Hearing systems or binaural hearing systems may further comprise one or more 'auxiliary devices', which communicate with the hearing aid(s) and affect and/or benefit from the function of the hearing aid(s). Auxiliary devices may be e.g. remote controls, audio gateway devices, mobile phones (e.g. Smartphones), public-address systems, car audio systems or music players. Hearing aids, hearing systems or binaural hearing systems may e.g. be used for compensating for a hearing-impaired person's loss of hearing capability, augmenting or protecting a normal-hearing person's hearing capability and/or conveying electronic audio signals to a person.

BRIEF DESCRIPTION OF DRAWINGS

[0069] The aspects of the disclosure may be best understood from the following detailed description taken in conjunction with the accompanying figures. The figures are schematic and simplified for clarity, and they just show details to improve the understanding of the claims, while other details are left out. Throughout, the same reference numerals are used for identical or corresponding parts. The individual features of each aspect may each be combined with any or all features of the other aspects. These and other aspects, features and/or technical effect will be apparent from and elucidated with reference to the illustrations described hereinafter in which:

FIG. 1A schematically shows a time variant analogue signal (Amplitude vs time) and its digitization in samples, the samples being arranged in a number of time frames, each comprising a number N_s of samples, and

FIG. 1B illustrates a time-frequency map representation of the time variant electric signal of FIG. 1A,

FIG. 2A symbolically shows an intrusive monaural speech intelligibility predictor unit providing a monaural speech intelligibility predictor d based on a time-frequency representation $x_j(m)$ of an information signal x ,

FIG. 2B shows a first embodiment an intrusive monaural speech intelligibility predictor unit, and

FIG. 2C shows a second embodiment an intrusive monaural speech intelligibility predictor unit.

FIG. 3A shows an intrusive monaural speech intelligibility predictor unit in combination with a hearing loss model and an evaluation unit,

FIG. 3B shows an intrusive monaural speech intelligibility predictor unit in combination with a signal processing unit and an evaluation unit, and

FIG. 3C shows a first embodiment of a hearing aid comprising an intrusive monaural speech intelligibility predictor unit with a hearing loss model configured to optimize a user's intelligibility of an output signal of the hearing aid,

FIG. 4A shows a first scenario for using a hearing aid comprising an intrusive monaural speech intelligibility predictor according to the present disclosure to improve a hearing aid user's intelligibility of speech from a speaker wearing a wireless microphone (e.g. in a teaching or lecture situation), and

FIG. 4B shows a second scenario for using a hearing aid comprising an intrusive monaural speech intelligibility predictor according to the present disclosure to improve a hearing aid user's intelligibility of speech from a remote speaker of a telephone conversation using a handsfree telephone set in a car, where remote sound is wirelessly as well as acoustically transmitted to the hearing aid user,

FIG. 5A shows a third scenario for using a hearing aid comprising an intrusive monaural speech intelligibility predictor according to the present disclosure to improve a hearing aid user's intelligibility of speech from a remote speaker of a telephone conversation using a handsfree telephone set in a car, where remote sound is (only) wirelessly transmitted to the hearing aid user,

FIG. 5B shows an embodiment of a hearing aid comprising an intrusive monaural speech intelligibility predictor for use in the (third) scenario of FIG. 5A, and

FIG. 5C illustrates an exemplary hearing aid formed as a receiver in the ear (RITE) type of hearing aid comprising a part adapted for being located behind pinna and a part comprising an output transducer (e.g. a loudspeaker/receiver) adapted for being located in an ear canal of the user,

FIG. 6A shows a binaural hearing aid system according to the present disclosure comprising first and second hearing aids and an auxiliary device, and

FIG. 6B shows the auxiliary device comprising a user interface in the form of an APP for controlling and displaying data related to the speech intelligibility predictors, and

FIG. 7 shows an embodiment of a binaural hearing aid system according to the present disclosure comprising a left and right hearing devices, each comprising a monaural speech intelligibility predictor whose individual predictor values d_{left} and d_{right} are exchanged between the hearing devices and used to influence or control signal processing in the hearing devices to optimize binaural speech intelligibility of the user.

[0070] The figures are schematic and simplified for clarity, and they just show details which are essential to the understanding of the disclosure, while other details are left out. Throughout, the same reference signs are used for identical or corresponding parts.

[0071] Further scope of applicability of the present disclosure will become apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating preferred embodiments of the disclosure, are given by way of illustration only. Other embodiments may become apparent to those skilled in the art from the following detailed description.

DETAILED DESCRIPTION OF EMBODIMENTS

[0072] The detailed description set forth below in connection with the appended drawings is intended as a description of various configurations. The detailed description includes specific details for the purpose of providing a thorough understanding of various concepts. However, it will be apparent to those skilled in the art that these concepts may be practised without these specific details. Several aspects of the apparatus and methods are described by various blocks, functional units, modules, components, circuits, steps, processes, algorithms, etc. (collectively referred to as "elements"). Depending upon particular application, design constraints or other reasons, these elements may be implemented using electronic hardware, computer program, or any combination thereof.

[0073] The electronic hardware may include microprocessors, microcontrollers, digital signal processors (DSPs), field programmable gate arrays (FPGAs), programmable logic devices (PLDs), gated logic, discrete hardware circuits, and other suitable hardware configured to perform the various functionality described throughout this disclosure. Computer program shall be construed broadly to mean instructions, instruction sets, code, code segments, program code, programs, subprograms, software modules, applications, software applications, software packages, routines, subroutines, objects, executables, threads of execution, procedures, functions, etc., whether referred to as software, firmware, middleware, microcode, hardware description language, or otherwise.

[0074] The present application relates to the field of hearing aids or hearing aid systems.

[0075] The present disclosure relates to signal processing methods for predicting the intelligibility of speech, e.g., the output signal of a signal processing device such as a hearing aid. The intelligibility prediction is made in the form of an index that correlates highly with the fraction of words that an average listener would be able to understand from some speech material. For situations where an estimate of absolute intelligibility, i.e., the actual percentage of words understood, is desired, this index may be transformed to a number in the range 0-100 percent, see e.g. [3] for one method to do this.

[0076] The method proposed here belongs to the class of so-called *intrusive* methods. Methods in this class are characterized by the fact that they make their intelligibility prediction by comparing the noisy - and potentially signal processed - speech signal, with a noise-free, undistorted version of the underlying speech signal, see [1, 2, 3] for examples of existing methods. The assumption that a noise-free reference signal is available is reasonable in many practically relevant situations. For example, when evaluating the impact of various hearing aid signal processing algorithms on intelligibility, one normally conducts a listening test with human subjects. In preparing such a test, the stimuli are often created artificially by explicitly adding noise signal to noise-free speech signals - in other words, noise-free signals are readily available. Hence, the proposed intelligibility prediction algorithm allows one to replace a costly and time-consuming listening test involving human subjects, with machine predictions.

[0077] Much of the signal processing of the present disclosure is performed in the time-frequency domain, where a time domain signal is transformed into the (time-)frequency domain by a suitable mathematical algorithm (e.g. a Fourier transform algorithm) or filter (e.g. a filter bank).

[0078] FIG. 1A schematically shows a time variant analogue signal (Amplitude vs time) and its digitization in samples, the samples being arranged in a number of time frames, each comprising a number N_s of digital samples. FIG. 1A shows an analogue electric signal (solid graph), e.g. representing an acoustic input signal, e.g. from a microphone, which is converted to a digital audio signal in an analogue-to-digital (AD) conversion process, where the analogue signal is sampled with a predefined sampling frequency or rate f_s , f_s being e.g. in the range from 8 kHz to 40 kHz (adapted to the particular needs of the application) to provide digital samples $x(n)$ at discrete points in time n , as indicated by the vertical lines extending from the time axis with solid dots at its endpoint coinciding with the graph, and representing its digital sample value at the corresponding distinct point in time n . Each (audio) sample $x(n)$ represents the value of the acoustic signal at n by a predefined number N_b of bits, N_b being e.g. in the range from 1 to 16 bits. A digital sample $x(n)$ has a length in time of $1/f_s$, e.g. 50 μ s, for $f_s = 20$ kHz. A number of (audio) samples N_s are arranged in a time frame, as schematically illustrated in the lower part of FIG. 1A, where the individual (here uniformly spaced) samples are grouped in time frames (1, 2, ..., N_s). As also illustrated in the lower part of FIG. 1A, the time frames may be arranged consecutively to be non-overlapping (time frames 1, 2, ..., m , ..., M) or overlapping (here 50%, time frames 1, 2, ..., m , ..., M'), where m is time frame index. In an embodiment, a time frame comprises 64 audio data samples. Other frame lengths may be

used depending on the practical application.

[0079] FIG. 1B schematically illustrates a time-frequency representation of the (digitized) time variant electric signal $x(n)$ of FIG. 1A. The time-frequency representation comprises an array or map of corresponding complex or real values of the signal in a particular time and frequency range. The time-frequency representation may e.g. be a result of a Fourier transformation converting the time variant input signal $x(n)$ to a (time variant) signal $x(k,m)$ in the time-frequency domain. In an embodiment, the Fourier transformation comprises a discrete Fourier transform algorithm (DFT). The frequency range considered by a typical hearing device (e.g. a hearing aid) from a minimum frequency $f_{\min}$ to a maximum frequency $f_{\max}$ comprises a part of the typical human audible frequency range from 20 Hz to 20 kHz, e.g. a part of the range from 20 Hz to 12 kHz. In FIG. 1B, the time-frequency representation $x(k,m)$ of signal $x(n)$ comprises complex values of magnitude and/or phase of the signal in a number of DFT-bins defined by indices (k,m) , where $k=1, \dots, K$ represents a number K of frequency values (cf. vertical k -axis in FIG. 1B) and $m=1, \dots, M$ (M') represents a number M (M') of time frames (cf. horizontal m -axis in FIG. 1B). A time frame is defined by a specific time index m and the corresponding K DFT-bins (cf. indication of *Time frame m* in FIG. 1B). A time frame m represents a frequency spectrum of signal x at time m . A DFT-bin (k,m) comprising a (real) or complex value $x(k,m)$ of the signal in question is illustrated in FIG. 1B by hatching of the corresponding field in the time-frequency map. Each value of the frequency index k corresponds to a frequency range Δf_k , as indicated in FIG. 1B by the vertical frequency axis f . Each value of the time index m represents a time frame. The time Δt_m spanned by consecutive time indices depend on the length of a time frame (e.g. 25 ms) and the degree of overlap between neighbouring time frames (cf. horizontal t -axis in FIG. 1B).

[0080] In the present application, a number J of (non-uniform) frequency sub-bands with sub-band indices $j=1, 2, \dots, J$ is defined, each sub-band comprising one or more DFT-bins (cf. vertical *Sub-band j* -axis in FIG. 1B). The j^{th} sub-band (indicated by *Sub-band $j(x_j(m))$* in the right part of FIG. 1B) comprises DFT-bins with lower and upper indices $k1(j)$ and $k2(j)$, respectively, defining lower and upper cut-off frequencies of the j^{th} sub-band, respectively. A specific time-frequency unit (j,m) is defined by a specific time index m and the DFT-bin indices $k1(j)$ - $k2(j)$, as indicated in FIG. 1B by the bold framing around the corresponding DFT-bins. A specific time-frequency unit (j,m) contains complex or real values of the j^{th} sub-band signal $x_j(m)$ at time m .

[0081] FIG. 2A symbolically shows an intrusive monaural speech intelligibility predictor unit (*MSIP*) providing a monaural speech intelligibility predictor d based on either

- time domain versions $s(n)$, $x(n)$ (n being a time (sample) index),
- time-frequency band representations $s(k,m)$, $x(k,m)$ (k being a frequency index, m being a time (frame) index), or
- sub-band representations $s_j(m)$, $x_j(m)$ (j being a frequency sub-band index)

of an essentially noise-free (clean) version of a target signal s and a noisy and/or processed version x of the target signal s comprising speech, respectively.

[0082] FIG. 2B shows a first embodiment an intrusive monaural speech intelligibility predictor unit (*MSIP*). The intrusive monaural speech intelligibility predictor unit (*MSIP*) is adapted for receiving a target signal comprising speech in an essentially noise-free version $s(n)$ and in a noisy and/or processed version $x(n)$, where n is a time index. The monaural speech intelligibility predictor unit is configured to provide as an output a final monaural speech intelligibility predictor value d indicative of a listener's (user's) perception of the noisy and/or processed version x of the target signal.

[0083] The monaural speech intelligibility predictor unit (*MSIP*) comprises a first input unit (*IU*) for providing a time-frequency representation $s(k,m)$ of said noise-free version s of the target signal from the time variant signal $s(n)$, and a second input unit (*IU*) for providing a time-frequency representation $x(k,m)$ of the noisy and/or processed version x of the target signal from the the time variant signal $x(n)$, k being a frequency bin index, $k=1, 2, \dots, K$, and m being a time index.

[0084] The monaural speech intelligibility predictor unit (*MSIP*) further comprises a first envelope extraction unit (*AEU*) for providing a time-frequency sub-band representation $s_j(m)$ of the noise-free version s of the target signal representing temporal envelopes, or functions thereof, of frequency sub-band signals $s_j(m)$ of said noise-free target signal from the time-frequency representation $s(k,m)$, and a second envelope extraction unit (*AEU*) for providing a time-frequency sub-band representation $x_j(m)$ of the noisy and/or processed version x of the target signal representing temporal envelopes, or functions thereof, of frequency sub-band signals $x_j(m)$ of said noisy and/or processed version of the target signal from the time-frequency representation $s(k,m)$, $j=1, 2, \dots, J$, and m being the time index.

[0085] The monaural speech intelligibility predictor unit (*MSIP*) further comprises a first time-frequency segment division unit (*SDU*) for dividing said time-frequency sub-band representation $s_j(m)$ of the noise-free version s of the target signal into time-frequency segments S_m corresponding to a number N of successive samples of the sub-band signals $s_j(m)$, and a second time-frequency segment division unit (*SDU*) for dividing the time-frequency sub-band representation $x_j(m)$ of the noisy and/or processed version x of the target signal into time-frequency segments X_m corresponding to a number N of successive samples of the sub-band signals $x_j(m)$.

[0086] The monaural speech intelligibility predictor unit (*MSIP*) further optionally comprises a first normalization and/or transformation unit (*N/TU*) adapted for providing normalized and/or transformed versions $\tilde{S}_m$ of the time-frequency segments S_m , and optionally a second normalization and/or transformation unit (*N/TU*) adapted for providing normalized and/or transformed versions $\tilde{X}_m$ of the time-frequency segments X_m .

[0087] The monaural speech intelligibility predictor unit (*MSIP*) further comprises an intermediate speech intelligibility calculation unit (*ISIU*) adapted for providing intermediate speech intelligibility coefficients d_m estimating an intelligibility of the time-frequency segment X_m , wherein the intermediate speech intelligibility coefficients d_m are based on the essentially noise-free, optionally normalized and/or transformed, time frequency segments S_m , $\tilde{S}_m$, and the noisy and/or processed, optionally normalized and/or transformed, time-frequency segments X_m , $\tilde{X}_m$.

[0088] The monaural speech intelligibility predictor unit (*MSIP*) further comprises a final monaural speech intelligibility calculation unit (*FSIU*) for calculating a final monaural speech intelligibility predictor d estimating an intelligibility of the noisy and/or processed version x of the target signal by combining, e.g. by averaging or applying a MIN or MAX-function, the intermediate speech intelligibility coefficients d_m , or a transformed version thereof, over time.

[0089] FIG. 2C shows a second embodiment an intrusive monaural speech intelligibility predictor unit (*MSIP*). The embodiment of FIG. 2C comprises the same functional units as described in connection with FIG. 2B. Additionally, it comprises a voice activity detector unit (*VAD*) for indicating whether or not or to what extent a given time-segment of the essentially noise-free version s ($s'(n)$ in FIG. 2C) and the noisy and/or processed version x ($x'(n)$ in FIG. 2C), respectively, of the target signal comprises or is estimated to comprise speech, and providing a voice activity control signal indicative thereof. In the embodiment of FIG. 2C, the voice activity detector unit (*VAD*) itself is configured to provide modified versions of the essentially noise-free version s and the noisy and/or processed version x , respectively, of the target signal comprising only time segments comprising speech or being estimated to comprise speech (in FIG. 2C denoted $s(n)$ and $x(n)$ respectively). Alternatively, the modified signals $s(n)$ and $x(n)$ may be created in respective separation units.

[0090] In order to simulate the potential decrease in intelligibility due to a hearing loss, an optional hearing loss model is included (cf. FIG. 3A, 3B, 3C). Several methods for simulating a hearing loss exist [4]. Typically, a hearing loss model builds (at least) on an audiogram containing frequency dependent hearing thresholds of a user (or representative of a type of hearing loss). The, perhaps, simplest hearing loss model consists of adding to the input signal $x(n)$ a statistically independent noise signal, which is spectrally shaped according to the audiogram of the listener [5].

[0091] The proposed monaural, intrusive speech intelligibility predictor may be decomposed into a number of sub-stages as illustrated in FIG. 2B and 2C and discussed above. Each sub-stage is described in more detail in the following.

Voice Activity Detection (VAD).

[0092] Speech intelligibility (SI) relates to regions of the input signal with speech activity -

silence regions do not contribute to SI. Hence, the first step is to detect voice activity regions in the input signals. Since the noise-free speech signal $s'(n)$ is available, voice activity is trivial. For example, in [3] the noise-free speech signal $s'(n)$ was divided into successive frames. Speech-active frames were then identified as the ones with a frame-energy no less than e.g. 40 dB of the frame with maximum energy. The speech inactive frames, i.e., the ones with energy less than e.g., 40 dB of the maximum frame energy, are then discarded from both signals, $x'(n)$ and $s'(n)$. Let us denote the input signals with speech activity by $x(n)$ and $s(n)$, respectively, where n is a discrete-time index. A voice activity detector is shown in FIG. 2C as unit VAD.

Frequency Decomposition (IU) and Envelope Extraction (AEU)

[0093] The first step is to perform a frequency decomposition (cf. input unit IU in FIG. 2C) of the signals $x(n)$ and $s(n)$. This may be achieved in many ways, e.g., using a short-time Fourier transform (STFT), a band-pass filterbank (e.g., a Gamma-tone filter bank), etc. Subsequently, the temporal envelopes of each sub-band signal are extracted (cf. unit AEU in FIG. 2C). This may, e.g., be achieved using a Hilbert transform, or by low-pass filtering the magnitude of complex-valued STFT signals, etc.

[0094] As an example, we describe in the following how the frequency decomposition and envelope extraction can be achieved using an STFT; the described procedure is similar to the one in [3]. Let us assume, as an example, that signals are sampled with a frequency of $f_s=10000$ Hz. First, a time-frequency representation is obtained by segmenting signals $x(n)$ and $s(n)$ into (e.g. 50%) overlapping, windowed frames (cf. e.g. FIG. 1A). Normally, a tapered window, e.g. a Hanning-window, is used. The window length may e.g. be 256 samples when the sample rate is 10000 Hz. Then, each frame is Fourier transformed using a fast Fourier transform (FFT) (potentially after appropriate zero-padding). The resulting DFT bins may be grouped in perceptually relevant sub-bands. For example, one could use one-third octave bands (e.g. as in [4]), but it should be clear that any other sub-band division can be used (for example, the grouping could be uniform, i.e., unrelated to perception in this respect, cf. FIG. 1B). In the case of one-third octave bands and a sampling rate of 10000 Hz, there are 15 bands which cover the frequency range 150-5000 Hz. Other numbers of bands and another frequency range can be used. We refer to the time-frequency tiles defined by these frames and sub-bands as time-frequency (TF) units (or STFT coefficients), cf. FIG. 1B. Applying this to the noisy/processed input signal $x(n)$ leads to (generally complex-valued) STFT coefficients $x(k,m)$, where k and m denote frequency and frame (time) indices, respectively. STFT coefficients $s(k,m)$ for the noise-free reference speech signal are found in a similar manner.

[0095] Temporal envelope signals may then be extracted as

$$x_j(m) = f \left(\sqrt{\sum_{k=k1(j)}^{k2(j)} |x(k,m)|^2} \right), j = 1, \dots, J, \text{ and } m = 1, \dots, M,$$

where $k1(j)$ and $k2(j)$ denote DFT bin indices corresponding to lower and higher cut-off

frequencies of the j^{th} sub-band, J is the number of sub-bands (e.g. 16), and M is the number of signal frames in the signal in question, and where the function $f(x)$ is included for generality. For example, for $f(x) = x$, we get the temporal envelope used in [4], with $f(x) = x^2$, we extract power envelopes, and with $f(x) = 2 \log x$, or $f(x) = x^\beta$, $0 < \beta < 2$, we can model the compressive non-linearity of the healthy cochlea, respectively. It should be clear that other reasonable choices for $f(x)$ exist. Temporal envelope signals $s_j(m)$ for the noise-free speech signal are found in a similar manner. The same choice of $f(x)$ may be used in both cases.

[0096] As mentioned, other envelope representations may be implemented, e.g., using a Gammatone filterbank, followed by a Hilbert envelope extractor, etc., and functions $f(x)$ may be applied to these envelopes in a similar manner as described above for STFT based envelopes. In any case, the result of this procedure is a time-frequency representation in terms of sub-band temporal envelopes, $x_j(m)$ and $s_j(m)$, where j is a sub-band index, and m is a time index.

Time-Frequency Segments (SDU)

[0097] Next, we divide the time-frequency representations $x_j(m)$ and $s_j(m)$ into segments, i.e., spectrograms corresponding to N successive samples of all sub-band signals. For example, the m^{th} segment for the noisy/processed signal is defined by the $J \times N$ matrix

$$X_m = \begin{bmatrix} x_1(m-N+1) & \cdots & x_1(m) \\ \vdots & & \vdots \\ x_J(m-N+1) & \cdots & x_J(m) \end{bmatrix}.$$

The corresponding segment S_m for the noise-free reference signal is found in an identical manner.

[0098] It should be understood that other versions of the time-segments could be used, e.g., segments, which have been shifted in time to operate on frame indices $m - N/2 + 1$ through $m + N/2$.

Normalizations and Transformation of Time-Frequency Segments (N/TU)

[0099] The rows and columns of each segment X_m and S_m may be normalized/transformed in various ways (below, we show the normalizations/transformations as applied to X_m ; they are applied to S_m in a completely analogously manner. The same normalization/transformation is applied to both X_m and S_m). In particular, we consider the following row (R) normalizations/transformations

R1) Normalization of rows to zero mean:

$$g_1(X) = X - \mu_x^r \mathbf{1}^T,$$

where

$$\mu_x^r$$

is a $J \times 1$ vector whose j^{th} entry is the mean of the j^{th} row of X (hence the superscript r in μ_x^r),

and where $\mathbf{1}$ denotes an $N \times 1$ vector of ones.

R2) Normalization of rows to unit-norm:

$$g_2(X) = D^r(X)X,$$

where

$$D^r(X) = \text{diag}\left(\left[1/\sqrt{X(1,:)X(1,:)^H} \quad \dots \quad 1/\sqrt{X(J,:)X(J,:)^H}\right]\right),$$

and where $\text{diag}(\cdot)$ is a diagonal matrix with the elements of the arguments on the main diagonal. Furthermore, $X(j,:)$ denotes the j^{th} row of X , such that $D^r(X)$ is a $J \times J$ diagonal matrix with the inverse norm of each row on the main diagonal, and zeroes elsewhere (the superscript H denotes Hermitian transposition). Pre-multiplication with $D^r(X)$ normalizes the rows of the resulting matrix to unit-norm.

R3) Fourier transformation applied to each row

$$g_3(X) = XF,$$

where F is an $N \times N$ Fourier matrix.

R4) Fourier transformation applied to each row followed by computing the magnitude of the resulting complex-valued elements

$$g_4(X) = |XF|,$$

where $|\cdot|$ computes the element-wise magnitudes.

R5) The identity operator

$$g_5(X) = X.$$

[0100] We consider the following column (C) normalizations

C1) Normalization of columns to zero mean:

$$h_1(X) = X - \mathbf{1}\mu_x^{cT},$$

where

$$\mu_x^c$$

is a $N \times 1$ vector whose i^{th} entry is the mean of the i^{th} row of X , and where $\mathbf{1}$ denotes an $J \times 1$ vector of ones.

C2) Normalization of columns to unit-norm:

$$h_2(X) = XD^c(X),$$

where

$$D^c(X) = \text{diag}\left(\left[1/\sqrt{X(:,1)^H X(:,1)} \quad \dots \quad 1/\sqrt{X(:,N)^H X(:,N)}\right]\right).$$

Here $X(:,n)$ denotes the n^{th} row of X , such that $D^c(X)$ is a diagonal $N \times N$ matrix with the inverse norm of each column on the main diagonal, and zeros elsewhere. Post-multiplication with

$D^c(X)$ normalizes the rows of the resulting matrix to unit-norm.

[0101] The row- ($R\#, \# = 1, 2, \dots, 5$) and column ($C\#, \# = 1, 2$) normalizations/transformations listed above may be combined in different ways. In a preferred embodiment, at least one of row normalizations/transformations $g_i(\cdot)$ ($i = 1, 2, \dots, 5$) and at least one of the column normalizations/transformations $h_j(\cdot)$ ($j = 1, 2$) is applied (in any order).

[0102] One combination of particular interest is where, first, the rows are normalized to zero-mean and unit-norm, followed by a similar mean and norm normalization of the columns. This particular combination may be written as

$$\tilde{X}_m = h_2(h_1(g_2(g_1(X_m))))$$

where X_m is the resulting row- and column normalized matrix.

[0103] Another transformation of interest is to compute the magnitude Fourier spectrum of each row of matrix X_m followed by mean- and norm- normalization of the resulting columns.

With the introduced notation, this may be written simply as

$$\tilde{X}_m = h_2(h_1(g_3(X_m)))$$

[0104] Other combinations of these normalizations/transformations may be of interest, e.g.,

$$\tilde{X}_m = g_2(g_1(h_2(h_1(X_m))))$$

(mean- and norm- standardization of the columns followed by mean- and norm-standardization of the rows), etc. As mentioned, a particular combination of row- and column-normalizations/transformations is chosen and applied to all segments X_m and S_m of the noisy/processed and noise-free signal, respectively.

Estimation of Intermediate Intelligibility Coefficients (ISIU)

[0105] The time-frequency segments S_m or the normalized/transformed time-frequency segments $\tilde{S}_m$ of the noise-free reference signal may now be used together with the corresponding noisy/processed segments $X_m, \tilde{X}_m$ to compute an intermediate intelligibility index d_m , reflecting the intelligibility of the noisy/processed signal segment $X_m, \tilde{X}_m$. To do so, let us first define the sample correlation coefficient $d(x,y)$ of the elements in two $K \times 1$ vectors x and y :

$$d(x,y) = \frac{\sum_{k=1}^K (x(k) - \mu_x)(y(k) - \mu_y)}{\sqrt{\sum_{k=1}^K (x(k) - \mu_x)^2 (y(k) - \mu_y)^2}}$$

where

$$\mu_x = \frac{1}{K} \sum_{k=1}^K x(k)$$

$$\mu_x = \frac{1}{K} \sum_{k=1}^K x(k)$$

and

$$\mu_y = \frac{1}{K} \sum_{k=1}^K y(k).$$

[0106] Several options exist for computing the intermediate intelligibility index d_m . In particular, d_m may be defined as

1. 1) the average sample correlation coefficient of the *columns* in $\tilde{S}_m$ and $\tilde{X}_m$, i.e.,

$$d_m = \frac{1}{N} \sum_{n=1}^N d(\tilde{S}_m(:,n), \tilde{X}_m(:,n)),$$

or

2. 2) the average sample correlation coefficient of the *rows* in

$$\tilde{S}_m$$

and $\tilde{X}_m$, i.e.,

$$d_m = \frac{1}{J} \sum_{j=1}^J d(\tilde{S}_m(j,:), \tilde{X}_m(j,:)),$$

or

3. 3) the sample correlation coefficient of *all* elements in $\tilde{S}_m$ and $\tilde{X}_m$, i.e.,

$$d_m = d(\tilde{S}_m(:), \tilde{X}_m(:)),$$

where we adopted the notation $\tilde{S}_m(:)$ and $\tilde{X}_m(:)$ to represent $NJ \times 1$ vectors formed by stacking the columns of the respective matrices.

Estimation of Final Intelligibility Coefficient (FSIU)

[0107] The final intelligibility coefficient d , which reflects the intelligibility of the noisy/processed input signal $x(n)$, is defined as the average of the intermediate intelligibility coefficients, potentially transformed via a function $u(d_m)$, across the duration of the speech-active parts of $x(n)$, i.e.,

$$d = \frac{1}{M} \sum_{m=1}^M u(d_m).$$

The function $u(d_m)$ could for example be

$$u(d_m) = \log\left(\frac{1}{1-d_m^2}\right),$$

to link the intermediate intelligibility coefficients to information measures, but it should be clear that other choices exist.

[0108] The "do-nothing" function $u(d_m) = d_m$ is also a possible choice (it has previously been used in the STOI algorithm [3]).

[0109] In the following, a noisy/reverberant speech signal $x(n)$ which potentially has been passed through a signal processing device, e.g. in a hearing aid, is considered. An algorithm is proposed, which can predict the average intelligibility of $x(n)$, as perceived by a group of listeners with similar hearing profiles, e.g. normal hearing or hearing impaired listeners. To achieve this, the proposed algorithm relies on the presence of the noise-free, undistorted underlying signal $s(n)$, see FIG. 3A. FIG. 3A shows an intrusive monaural speech intelligibility predictor unit in combination with a hearing loss model (*HLM*) and an evaluation unit (*MSIP*) (together constituting a modified monaural speech intelligibility predictor unit (*MSIP'*)). In the embodiment of FIG. 3A, the signal $x'(n)$ is passed through hearing loss model (*HLM*) configured to model the imperfections of an impaired auditory system (e.g. the impaired auditory system of a particular user). The hearing loss model unit (*HLM*) is e.g. based on an audiogram of an ear of a user (and possible other data related to a user's hearing ability). The hearing loss model unit (*HLM*) is e.g. configured to apply a frequency dependent modification of the noisy and/or processed version x' of the target signal reflecting a deviation from normal hearing, e.g. a hearing impairment, of a relevant ear of the user to provide a modified noisy and/or processed version x of the target signal for use together with the essentially noise-free version s of the target signal as inputs to the monaural speech intelligibility predictor unit (*MSIP*) providing the final monaural speech intelligibility predictor d . In FIG. 3A an evaluation unit (*EVAL*) is shown to receive and evaluate the speech intelligibility predictor d and provide a processed predictor d' . The evaluation unit (*EVAL*) may e.g. further process the speech intelligibility predictor value d , to e.g. graphically and/or numerically display the current and/or recent historic values, derive trends, etc. Alternatively, or additionally the evaluation unit may propose actions to the user (or a communication partner or caring person), such as add directionality, move closer, speak louder, activate SI-enhancement mode, etc. The evaluation unit may e.g. be implemented in a separate device, e.g. acting as a user interface to the speech intelligibility predictor unit (*MSIP*) and/or to a hearing aid including such unit., e.g. implemented as a remote control device, e.g. as an APP of a smartphone (cf. FIG. 6A, 6B).

[0110] FIG. 3B shows an intrusive monaural speech intelligibility predictor unit (*MSIP'*) in combination with a signal processing unit (*SPU*) and an evaluation unit (*EVAL*), e.g. of a hearing device. Compared to FIG. 3A, the embodiment of FIG. 3B additionally comprises a number of input units (here 2: *M1*, *M2*), e.g. microphones, for providing a time-variant electric input signal representing a sound input received at the input unit in question. At least one (such as each) of the electric input signals comprises a target signal component (e.g. a speech component) and a noise signal component (termed *Noisy target* in FIG. 3B). The target signal component is assumed to originate from a target signal source in the environment of the device (e.g. a hearing device, see FIG. 3C). The embodiment of FIG. 3B further comprises a configurable signal processing unit (*SPU*) for processing the electric input signals (e.g. providing beamforming and/or noise reduction, frequency and level dependent amplification, level dependent compression, or the like) and providing a processed signal x' based on one or more of the electric input signals, which are inputs to the configurable signal processing unit

(*SPU*). The processed signal x' from the configurable signal processing unit (*SPU*) is fed to the hearing loss model (*HLM*) unit of the monaural speech intelligibility predictor unit (*MSIP'*). The hearing loss model unit (*HLM*) comprises a model of a hearing loss of a user (e.g. the user of the device) and is configured to shape an input signal to provide an output signal x representing a processed (possibly hearing loss compensated) and (again) deteriorated signal, which is fed to the monaural speech intelligibility predictor (*MSIP*). The embodiment of FIG. 3B further comprises an antenna and transceiver unit (*Rx*) for receiving a wireless signal (termed *Clean target* in FIG. 3B) comprising the target signal and for extracting an essentially noise-free version s of the target signal, which is connected to the monaural speech intelligibility predictor (*MSIP*). The final speech intelligibility predictor d from the monaural speech intelligibility predictor unit (*MSIP'*) is fed to the evaluation unit (*EVAL*) whose modified predictor value d' is fed to the configurable signal processing unit (*SPU*). The configurable signal processing unit (*SPU*) is adapted to control or influence the processing of the respective electric input signals based on the final speech intelligibility predictor d provided by the monaural speech intelligibility predictor unit and as modified by the evaluation unit (*EVAL*). The configurable signal processing unit (*SPU*) is adapted to control or influence the processing of the respective electric input signals to maximize the final speech intelligibility predictor d . (e.g. controlled by the evaluation unit (*EVAL* (*max*))).

[0111] The embodiment of FIG. 3B may e.g. further comprise an output unit for creating output stimuli configured to be perceivable by the user as sound based on an electric output either in the form of the processed signal x' from the signal processing unit or a signal derived therefrom. The output unit (cf. e.g. *OT* in FIG. 3C) may e.g. comprise a loudspeaker for placement in an ear canal of a user, or a vibrator for being attached to the skull of a user, or electrodes for placement in cochlea of a user. Thereby a hearing aid according to the present disclosure is provided. The hearing aid may take the form of an air conducting hearing instrument, a bone-conducting hearing instrument, a cochlear implant prosthesis, an active ear-protection device, a headset, an earphone with active noise cancellation, etc.

[0112] FIG. 3C shows a first embodiment of a hearing device (*HD*, e.g. a hearing aid) comprising an intrusive monaural speech intelligibility predictor unit (*MSIP'*) comprising a hearing loss model part (*HLM*) and a predictor part (*MSIP*) configured to optimize a user's intelligibility (represented by index d) of an output signal u of the hearing device (*HD*). The embodiment of FIG. 3C is equivalent to the embodiment of FIG. 3B but further comprises an output unit comprising an output transducer (*OT*) in the form of a loudspeaker, which is directly connected to the output u of the signal processing unit (*SPU*). Further, the hearing aid of FIG. 3C only comprises one the input unit (*IT*) comprising a microphone for picking up a noisy representation y' of the target signal hearing aid and converting it to an electric input signal y , which is fed to the configurable signal processing unit (*SPU*). As in FIG. 3B, the antenna and transceiver unit (*Rx*) is adapted for receiving a wireless signal (termed s' in FIG. 3C) comprising the target signal and for extracting an essentially noise-free version s of the target signal, which is fed to the predictor part (*MSIP*) of the intrusive monaural speech intelligibility predictor unit (*MSIP'*). The monaural speech intelligibility predictor unit (*MSIP'*) provides an estimate of the intelligibility of the output signal by the user in the form of the (final) speech

intelligibility predictor d , which is fed to a control part of the configurable signal processing unit (SPU) to modify signal processing to optimize d in a feedback loop.

[0113] FIG. 4A shows a first scenario for using a hearing aid (HD) comprising an intrusive monaural speech intelligibility predictor according to the present disclosure (as described in connection with FIG. 2A, 2B, 2C and FIG. 3A, 3B, 3C above) to improve a hearing aid user's (U) intelligibility of speech from a speaker (TLK) wearing a wireless microphone (M), e.g. in a teaching or lecture situation. The speaker's voice (the target signal) is picked up by the microphone (M) located close to the speaker's mouth. The microphone (M) comprises a transmitter (Tx) for wirelessly transmitting the essentially noise-free version of the target signal s to a corresponding receiver (Rx) of the hearing aid worn by the user (U) using wireless link WLS (e.g. using FM or Bluetooth or other standardized or proprietary technology). Simultaneously, an acoustically propagated version of the target signal coloured (modified) by the location (e.g. a room with reflecting surfaces, e.g. walls) and mixed with possible noise (noise) from the environment is picked up (noisy signal x) by one or more microphone of the hearing aid (HD).

[0114] FIG. 4B shows a second (similar) scenario for using a hearing aid (HD) comprising an intrusive monaural speech intelligibility predictor according to the present disclosure to improve a hearing aid user's (U) intelligibility of speech from a remote speaker of a telephone conversation using a handsfree telephone set in a car (CAR), where remote sound is wirelessly as well as acoustically transmitted to the hearing aid user. The 'clean' target signal of the remote speaker received by the telephone ($CELL PHONE$) is wirelessly transmitted to and received by a receiver (Rx) of the hearing aid (HD) as clean signal s (*Clean target s* in FIG. 4B), while the sound is simultaneously propagated to a microphone (IT) of the hearing aid by a loudspeaker ($LOUDSPEAKER$) of the handsfree telephone set providing 'noisy signal' x (*Noisy target x* in FIG. 4B). The user is driving a car (CAR , $WHEEL$) while talking in a telephone via a handsfree telephone set. The noise in the car cabin (indicated by noise sources $N1$, $N2$) is acoustically mixed with the 'target signal' played by a loudspeaker of the handsfree telephone set (or of the telephone itself) and picked up as noisy target signal x by the hearing aid microphone(s) (IT).

[0115] The hearing aid (HD) used in the two scenarios of FIG. 4A, 4B may be a hearing aid according to the present disclosure, e.g. as described in connection with FIG. 3A, 3B, 3C, which is configured to adapt the processing of an acoustic signal picked up by a microphone of the hearing aid and processed by a signal processing device to optimize the user's speech intelligibility (based on a predictor of a monaural speech intelligibility predictor unit, as proposed by the present disclosure).

[0116] FIG. 5A shows a third scenario for using a hearing aid (HD) comprising an intrusive monaural speech intelligibility predictor according to the present disclosure to improve a hearing aid user's intelligibility of speech from a remote speaker of a telephone conversation using a handsfree telephone set in a car (CAR , $WHEEL$), where sound from a remote communication partner is wirelessly transmitted to the hearing aid user (U). The scenario of

FIG. 5A is similar to the scenario of FIG. 4B apart from the fact that in the scenario of FIG. 5A the voice of the remote communication partner is NOT played by a loudspeaker in the car.

[0117] The clean target signal s is transmitted from the *CELL PHONE* to the hearing aid *HD*. The background noise v' (*Noise v'*) of the car cabin is captured by the microphone(s) (*IT*) of the hearing aid. It can be assumed that the background noise v' as captured is substantially equal to the noise v_{ed} (*Noise v_{ed}*) that is present at the ear drum (*Ear drum*) of the user (cf. FIG. 5B, 5C). The assumption is of course better the closer to the ear drum the microphone is situated (and/or the more open the ear canal part is). In an embodiment, a microphone of the hearing aid is located in the ear canal, e.g. at the entrance of the ear canal or close to the ear drum (cf. e.g. IT_3 in FIG. 5C).

[0118] FIG. 5B shows an embodiment of a hearing aid comprising an intrusive monaural speech intelligibility predictor for use in the (third) scenario of FIG. 5A. The embodiment of a hearing aid (*HD*) shown in FIG. 5B comprises the same functional components as the embodiment shown in FIG. 3C. The interconnection of the signal processing unit (*SPU*) and the wireless receiver (*Rx*) and the input transducer (*IT*) is different, however. As shown in FIG. 5A and as indicated in FIG. 5B and 5C, the sound output of the loudspeaker (*OT*) of the hearing aid, equal to the processed signal $f(s)$ from the signal processing unit (*SPU*), is acoustically mixed with 'environmental' (car cabin) noise v_{ed} at the eardrum (*Ear drum*, cf. *Mixture of s and v_{ed}* in FIG. 5B).

[0119] The basic idea of the embodiment of a hearing aid in FIG. 5B is to process the clean version s of the target signal so that the speech intelligibility d is maximized when the processed version of the clean target signal ($f(s)$). The processed version of the clean target signal ($f(s)$) can be adaptively controlled, whereas this is not the case for the car cabin noise v_{ed} at the eardrum (which is given).

[0120] Using a model of speech intelligibility (e.g. as disclosed in the present disclosure) in the configuration of FIG. 5B, an estimate of a user's intelligibility of speech components present in the processed signal $f(s)$ from the signal processing unit (*SPU*), when presented to the ear by a loudspeaker (or an acoustic guide element of the hearing aid) and mixed with the noise signal v_{ed} present at the ear drum, can be determined, cf. signal $u=f(s)+v$, which is used as the 'processed/noisy target signal input' to the hearing loss model (*HLM*) of the monaural speech intelligibility predictor unit (*MSIP'*) in FIG. 5B.

[0121] Preferably, the loudspeaker (or alternatively an acoustic guide element) is located in the ear canal, preferably close to the ear drum to deliver the processed signal $f(s)$ to the ear drum.

[0122] Preferably, the microphone(s) of the hearing device, which is(are) used to pick up background noise v' (cf. FIG. 5A, 5B), is(are) located close to the ear drum, or at the entrance of the ear canal, or in pinna, or behind the ear. In an embodiment, the noise signal picked up by a given microphone is appropriately modified (in the hearing aid) to account for a transfer

function for an acoustic signal from the location of the microphone to the ear drum to thereby give a better estimate of the background noise v_{ed} received by the user (at the ear drum) for use in the estimate of the speech intelligibility $d=d(f(s)+v,s)$.

[0123] In the configuration of FIG. 5B, the signal processing unit (*SPU*) is configured to iteratively modify signal processing of the clean target signal s received from wireless receiver unit (*Rx*) to provide processed version $f(s)$ the clean target signal s that optimizes speech intelligibility of the (mixed) signal present at the ear drum of the user (in practice here approximated by maximizing the monaural speech intelligibility predictor $d(f(s)+v,s)$ according to the resent disclosure).

[0124] As an alternative to using a speech intelligibility predictor to modify (optimize) s (or as an extreme option of the present disclosure), a simple increase of gain of the clean target signal s (i.e. $f(s)=g \cdot s$, g being a gain factor, e.g. $g=10$) may be used to increase the signal to noise ratio (SNR) at the ear drum (assuming a constant level of the background (cabin) noise v_{ed} at the ear drum). In practice, such reliance *only* on increasing gain of the clean target signal may, however, not be attractive or possible (e.g. due to acoustic feedback problems, maximum power output limitations of the loudspeaker, or uncomfortable levels of the user, etc.). Instead an appropriate frequency dependent shaping of the clean target signal is generally proposed and governed by the monaural speech intelligibility predictor (including the hearing loss model (*HLM*) preferably defining decisive aspects of a hearing impairment of the user of the hearing aid).

[0125] FIG. 5C illustrates an exemplary hearing aid (*HD*) formed as a receiver in the ear (*RITE*) type of hearing aid comprising a part (*BTE*) adapted for being located behind pinna and a part (*ITE*) comprising an output transducer (OT, e.g. a loudspeaker/receiver) adapted for being located in an ear canal (*Ear canal*) of the user (e.g. exemplifying a hearing aid (*HD*) as shown in FIG. 5A, 5B). The BTE-part (*BTE*) and the ITE-part (*ITE*) are connected (e.g. electrically connected) by a connecting element (*IC*). In the embodiment of a hearing aid of FIG. 5C, the BTE part (*BTE*) comprises two input units comprising two (individually selectable) input transducers (e.g. microphones) (IT_1 , IT_2) each for providing an electric input audio signal representative of an input sound signal from the environment (in the scenario of FIG. 5A, from the car cabin). The hearing device of FIG. 5C further comprises two (individually selectable) wireless receivers (WLR_1 , WLR_2) for providing respective directly received auxiliary audio and/or information signals. The hearing aid (*HD*) further comprises a substrate (*SUB*) whereon a number of electronic components are mounted, including a configurable signal processing unit (*SPU*), a monaural speech intelligibility predictor unit (*MSIP*), and a hearing loss model unit (*HLM*, coupled to each other and input and output units via electrical conductors Wx). The configurable signal processing unit (*SPU*) provides an enhanced audio signal (cf. signal $f(s)$ in FIG. 5B), which is intended to be presented to a user. In the embodiment of a hearing aid device in FIG. 5C, the ITE part (*ITE*) comprises an output unit in the form of a loudspeaker (receiver) (OT) for converting an electric signal ($f(s)$ in FIG. 5B) to an acoustic signal. The ITE-part further comprises an input unit comprising an input transducer (e.g. a microphone) (IT_3)

for providing an electric input audio signal representative of an input sound signal from the environment in the ear canal (here approximating the noise v_{ed} from the car cabin at the ear drum (*Ear drum*) of the user (*U*) wearing the hearing aid (*HD*)). In other embodiments, the hearing aid may comprise *only* the input unit (*IT*₃) located in or at the ear canal, or the input unit (*IT*₃) located in or at the ear canal in combination with a an input unit located elsewhere, e.g. in a BTE-part. The ITE-part further comprises a guiding element, e.g. a dome, (*DO*) for guiding and positioning the ITE-part in the ear canal of the user.

[0126] The hearing aid (*HD*) exemplified in FIG. 5C is a portable device and further comprises a battery (*BAT*) for energizing electronic components of the BTE- and ITE-parts.

[0127] In an embodiment, the hearing aid (*HD*) comprises a directional microphone system (beamformer) adapted to enhance a target acoustic source among a multitude of acoustic sources in the local environment of the user wearing the hearing aid device. In an embodiment, the directional system is adapted to detect (such as adaptively detect) from which direction a particular part of the microphone signal originates.

[0128] The hearing aid of FIG. 5C may form part of a hearing aid and/or a binaural hearing aid system according to the present disclosure (cf. e.g. FIG. 7).

[0129] FIG. 6A shows an embodiment of a binaural hearing system comprising left and right hearing devices (*HD*_{left}, *HD*_{right}), e.g. hearing aids, in communication with a portable (handheld) auxiliary device (*Aux*) functioning as a user interface (*UI*) for the binaural hearing aid system (cf. FIG. 6B). In an embodiment, the binaural hearing system comprises the auxiliary device (*Aux*, and the user interface *UI*). In the embodiment of FIG. 6A, wireless links denoted *IA-WL* (e.g. an inductive link between the left and right hearing devices) and *WL-RF* (e.g. RF-links (e.g. Bluetooth) between the auxiliary device *Aux* and the left *HD*_{left}, and between the auxiliary device *Aux* and the right *HD*_{right} hearing aid, respectively) are indicated (implemented in the devices by corresponding antenna and transceiver circuitry, indicated in FIG. 6A in the left and right hearing devices as *RF-IA-Rx/Tx-l* and *RF-IA-Rx/Tx-r*, respectively).

[0130] FIG. 6B shows the auxiliary device (*Aux*) comprising a user interface (*UI*) in the form of an APP for controlling and displaying data related to the speech intelligibility predictors. The user interface (*UI*) comprises a display (e.g. a touch sensitive display) displaying a screen of a *Speech intelligibility SI-APP* for controlling the hearing aid system and presenting information to the user. The APP comprises a number of predefined action buttons regarding functionality of the binaural (or monaural) hearing system. In the exemplified (part of the) APP, a user (*U*) has the option of influencing a mode of operation via the selection of a *SI-prediction mode* to be a *Monaural SIP* or *Binaural SIP* mode. In the screen shown in FIG. 10B, the un-shaded buttons are selected, i.e. *SI-enhancement mode* (where the processing is adapted to optimize speech intelligibility based on the (monaural or binaural) speech intelligibility predictor) together with a specific *Car telephony mode* (as described in connection with FIG. 5A, 5B, 5C). Further, a show *Current SI-estimate* has been activated (resulting in a current predicted value of the

binaural speech intelligibility predictor being displayed (in the form of the positive indicator '  ')

together with an indication of the current noise level (indicated as 'HIGH')). The grey shaded button *Lecture mode* (as described in connection with FIG. 4A, 4B) may be selected instead of *Car telephony mode*.

[0131] FIG. 7 shows an embodiment of a binaural hearing aid system according to the present disclosure comprising a left and right hearing devices (HD_{left} , HD_{right}), each comprising a monaural speech intelligibility predictor unit ($MSIP'$) whose individual predictor values d_{left} and d_{right} are exchanged between the hearing devices and used to influence or control signal processing of respective signal processing units (SPU) in the hearing devices to optimize binaural speech intelligibility of the user. The left and right hearing devices (HD_{left} , HD_{right}), are e.g. hearing devices as shown in an discussed in connection with FIG. 3C. Each of the left and right hearing aids comprises antenna and transceiver circuitry ($IA-Rx/Tx$) for allowing a communication link ($IA-WLS$) to be established and information (d_{left} , d_{right}) to be exchanged between said left and right hearing aids. In an embodiment, the binaural hearing aid system comprises a binaural speech intelligibility prediction unit for providing a final binaural speech intelligibility measure $d_{binaural}$ of the predicted speech intelligibility of the user when exposed to a sound input, wherein the final binaural speech intelligibility measure $d_{binaural}$ is determined in dependence of the final monaural speech intelligibility predictor values d_{left} , d_{right} of the respective left and right hearing aids. The binaural speech intelligibility prediction unit may e.g. be implemented in one or both of the signal processing units (SPU) of the left and right hearing devices. In an embodiment, the final binaural speech intelligibility measure $d_{binaural}$ is determined as the maximum of the speech intelligibility predictor values d_{left} , d_{right} of the respective left and right hearing aids: $d_{binaural} = \max(d_{left}, d_{right})$.

[0132] It is intended that the structural features of the devices described above, either in the detailed description and/or in the claims, may be combined with steps of the method, when appropriately substituted by a corresponding process.

[0133] As used, the singular forms "a," "an," and "the" are intended to include the plural forms as well (i.e. to have the meaning "at least one"), unless expressly stated otherwise. It will be further understood that the terms "includes," "comprises," "including," and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. It will also be understood that when an element is referred to as being "connected" or "coupled" to another element, it can be directly connected or coupled to the other element but an intervening elements may also be present, unless expressly stated otherwise. Furthermore, "connected" or "coupled" as used herein may include wirelessly connected or coupled. As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items. The steps of any disclosed method is not limited to the exact order stated herein, unless expressly stated otherwise.

[0134] It should be appreciated that reference throughout this specification to "one embodiment" or "an embodiment" or "an aspect" or features included as "may" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the disclosure. Furthermore, the particular features, structures or characteristics may be combined as suitable in one or more embodiments of the disclosure. The previous description is provided to enable any person skilled in the art to practice the various aspects described herein. Various modifications to these aspects will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other aspects.

[0135] The claims are not intended to be limited to the aspects shown herein, but is to be accorded the full scope consistent with the language of the claims, wherein reference to an element in the singular is not intended to mean "one and only one" unless specifically so stated, but rather "one or more." Unless specifically stated otherwise, the term "some" refers to one or more.

[0136] Accordingly, the scope should be judged in terms of the claims that follow.

REFERENCES

[0137]

[1] American National Standards Institute, "ANSI S3.5, Methods for the Calculation of the Speech Intelligibility Index," New York 1995.

[2] K. S. Rhebergen and N. J. Versfeld, "A speech intelligibility index based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am., vol. 117, no. 4, pp. 2181-2192, 2005.

[3] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no.7, pp. 2125-2136, Sept. 2011.

[4] B. C. J. Moore, "Cochlear Hearing Loss," Physiological, Psychological and Technical Issues, " Wiley, 2007.

[5] R. Beutelmann and T. Brand, "Prediction of intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am., Vol. 120, no. 1, pp. 331-342, April 2006.

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **TAGHIA et al.** On Mutual Information as a Measure of Speech Intelligibility Proc. ICASSP, 2012, [0002]
- **JENSENTAAL** Prediction of Intelligibility of Noisy and Time-Frequency Weighted Speech based on Mutual Information Between Amplitude Envelopes Proc. Interspeech, 2013, [0003]
- **TANG YAN et al.** Evaluating the predictions of objective intelligibility metrics for modified and synthetic speech COMPUTER SPEECH AND LANGUAGE, 2016, vol. 35, 73-92 [0004]
- **FEI CHEN** The relative importance of temporal envelope information for intelligibility prediction: A study on cochlear-implant vocoded speech MEDICAL ENGINEERING & PHYSICS, 2011, vol. 33, 81033-1038 [0005]
- **ANSI S3.5**, Methods for the Calculation of the Speech Intelligibility Index American National Standards Institute 19950000 [0137]
- **K. S. RHEBERGENN. J. VERSFELDA** A speech intelligibility index based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners J. Acoust. Soc. Am., 2005, vol. 117, 42181-2192 [0137]
- **C. H. TAALR. C. HENDRIKSR. HEUSDENSJ. JENSEN** An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech IEEE Trans. Audio, Speech, Lang. Process., 2011, vol. 19, 72125-2136 [0137]
- **Cochlear Hearing LossB. C. J. MOORE** Physiological, Psychological and Technical Issues Wiley 20070000 [0137]
- **R. BEUTELMANNT. BRAND** Prediction of intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners J. Acoust. Soc. Am., 2006, vol. 120, 1331-342 [0137]

Patentkrav

1. Monaural taleforståelighedsforudsigelsesenhed (MSIP), der er indrettet til at modtage et målsignal, som omfatter tale i en i det væsentlige støjfri version s og i en støjende og/eller behandlet version x , hvor den monaurale taleforståelighedsforudsigelsesenhed er konfigureret til at tilvejebringe, som et output, en endelig taleforståelighedsforudsigelsesværdi d , der er indikativ for en lytters opfattelse af den støjende og/eller behandlede version x af målsignalet, hvor den monaurale taleforståelighedsforudsigelsesenhed (MSIP) omfatter

- en første indgangsenhed (IU) til tilvejebringelse af en tids-frekvens-repræsentation $s(k,m)$ for den støjfri version s af målsignalet, hvor k er et frekvensområdeindeks, $k=1, 2, \dots, K$, og m er et tidsindeks;
- en anden indgangsenhed (IU) til tilvejebringelse af en tids-frekvens-repræsentation $x(k,m)$ af den støjende og/eller behandlede version x af målsignalet, k er et frekvensområdeindeks, $k=1, 2, \dots, K$, og m er et tidsindeks;

- en første indhyllings-ekstraktionsenhed (AEU) til tilvejebringelse af en tids-frekvens-subbånds-repræsentation $s_j(m)$ for den støjfri version s af målsignalet, der repræsenterer temporale indhyllinger, eller funktioner deraf, af frekvens-subbånds-signaler $s_j(m)$ for det støjfri målsignal, j er et frekvens-subbånds-indeks, $j=1, 2, \dots, J$, og m er tidsindekset;

- en anden indhyllings-ekstraktionsenhed (AEU) til tilvejebringelse af en tids-frekvens-subbånds-repræsentation $x_j(m)$ for den støjende og/eller behandlede version x af målsignalet, der repræsenterer temporale indhyllinger, eller funktioner deraf, af frekvens-subbånds-signaler $x_j(m)$ for den støjende og/eller behandlede version af målsignalet, $j=1, 2, \dots, J$, og m er tidsindekset;

- en første tids-frekvens-segmentdelingsenhed (SDU) til deling af tids-frekvens-subbånds-repræsentationen $s_j(m)$ for den støjfri version s af målsignalet til tids-frekvens-segmenter S_m , der svarer til et antal N af på hinanden følgende prøver af subbånds-signalerne;

- en anden tids-frekvens-segmentdelingsenhed (SDU) til deling af tids-frekvens-subbånds-repræsentationen $x_j(m)$ for den støjende og/eller behandlede version x af målsignalet til tids-frekvens-segmenter X_m , der svarer til et antal N af på hinanden følgende prøver af subbånds-signalerne;

hvor den første og anden tids-frekvens-segmentdelingsenhed er konfigureret til at dele tids-frekvens-repræsentationerne henholdsvis $s_j(m)$ og $x_j(m)$ i segmenter i form af spektrogrammer, der svarer til N på hinanden følgende prøver af alle subbånds-signaler, hvor det m^{te} segment er defineret af $J \times N$ matrixen

$$Z_m = \begin{bmatrix} z_1(m-N+1) & \cdots & z_1(m) \\ \vdots & & \vdots \\ z_J(m-N+1) & \cdots & z_J(m) \end{bmatrix},$$

hvor z , Z repræsenterer henholdsvis s , S eller x , X ;

- en normaliserings- og transformationsenhed (N/TU), der er konfigureret til at tilvejebringe i det mindste en normaliserings- og/eller transformationsfunktion af rækker og i det mindste en normaliserings- og/eller transformationsfunktion af kolonner af tids-frekvens-segmenterne S_m and X_m ;
- en mellemliggende taleforståelighedsberegningseenhed (ISIU), der er indrettet til tilvejebringelse af mellemliggende taleforståelighedskoefficienter d_m , der estimerer en forståelighed af tids-frekvens-segmentet X_m , hvor de mellemliggende taleforståelighedskoefficienter d_m er baseret på de i det væsentlige støjfrie, normaliserede og/eller transformerede tids-frekvens-segmenter $\tilde{S}_m$, og de støjende og/eller behandlede, normaliserede og/eller transformerede tids-frekvens-segmenter $\tilde{X}_m$;
- en endelig monaural taleforståelighedsberegningseenhed (FSIU) til beregning af en endelig taleforståelighedsforudsigelse d , der estimerer en forståelighed af den støjende og/eller behandlede version x af målsignalet ved at kombinere, for eksempel ved gennemsnit eller ved at anvende en MIN eller MAX-funktion til, de mellemliggende taleforståelighedskoefficienter d_m , eller en transformeret version deraf, over tid.

2. Monaural taleforståelighedsforudsigelseseenhed (MSIP) ifølge krav 1, der omfatter en stemmeaktivitetsdetektorenhed til at indikere, hvorvidt eller ikke eller i hvilken udstrækning et givet tidssegment af henholdsvis den i det væsentlige støjfrie version s og den støj- og/eller behandlede version x af målsignalet omfatter eller er estimeret til at omfatte tale, og tilvejebringe et stemmeaktivitetsstyresignal der er indikativ deraf.

3. Monaural taleforståelighedsforudsigelseseenhed (MSIP) ifølge krav 1 eller 2, der omfatter en stemmeaktivitetsdetektorenhed til identifikation af tidssegmenter for henholdsvis den i det væsentlige støjfrie version s og den støj- og/eller behandlede version x af målsignalet, der omfatter eller estimeres til at omfatte tale, og hvor den monaurale taleforståelighedsforudsigelseseenhed er konfigureret til at tilvejebringe modificerede versioner af henholdsvis den i det væsentlige støjfrie version s og for den støjende og/eller behandlede version x af målsignalet, hvor de modificerede versioner kun omfatter sådanne tidssegmenter, som omfatter tale eller er estimeret til at omfatte tale.

4. Monaural taleforståelighedsforudsigelseseenhed (MSIP) ifølge et hvilket som helst af kravene 1-3, der omfatter en høretabsmodelenhed, som er konfigureret til at anvende en modifikation af den støj- og/eller behandlede version x af målsignalet, der afspejler en afvigelse fra normal høreelse af et relevant øre for lytteren for at tilvejebringe en modificeret støj- og/eller behandlet version x af målsignalet til anvendelse sammen med den i det væsentlige støjfrie version s af målsignalet som en basis til beregning af den endelige taleforståelighedsforudsigelse d .

5. Monaural taleforståelighedsforudsigelsesenhed (HLM) ifølge et hvilket som helst af kravene 1-4, hvor høretabsmodelenheden er konfigureret til at tilføje et statistisk uafhængigt støjsignal, som er spektralt formet i overensstemmelse med et audiogram for lytterens relevante øre, til den støjende og/eller behandlede version x af målsignalet.

6. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge et hvilket som helst af kravene 1-5, der er indrettet til at ekstrahere de temporale indhyllingssignaler henholdsvis $x_j(m)$ og $s_j(m)$ som

$$z_j(m) = f \left(\sqrt{\sum_{k=k1(j)}^{k2(j)} |z(k, m)|^2} \right),$$

hvor z repræsenterer x eller s , $j=1, \dots, J$ og $m=1, \dots, M$, $k1(j)$ og $k2(j)$ betegner DFT områdeindekser, der svarer til lavere og højere grænsefrekvenser af j^{te} subbåndet, J er antallet af subbånd, og M er antallet af signalrammer i det pågældende signal, og $f(\cdot)$ er en funktion.

7. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge krav 6, hvor funktionen $f(\cdot)=f(w)$, hvor w repræsenterer

$$\left(\sqrt{\sum_{k=k1(j)}^{k2(j)} |z(k, m)|^2} \right),$$

er udvalgt blandt de følgende funktioner

- $f(w)=w$ repræsenterer identiteten
- $f(w)=w^2$ tilvejebringer effektindhyllinger,
- $f(w)=2 \cdot \log w$ eller $f(w)=w^\beta$, $0 < \beta < 2$, hvilket muliggør modellering af den komprimerende ulinearitet af sunde cochlea, eller kombinationer deraf.

8. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge et hvilket som helst af kravene 1-7, der omfatter

- en første normaliserings- og/eller transformationsenhed, der er indrettet til tilvejebringelse af normaliserede og/eller transformerede versioner $\tilde{S}_m$ af tids-frekvens-segmenterne S_m ;
- en anden normaliserings- og/eller transformationsenhed, der er indrettet til tilvejebringelse af normaliserede og/eller transformerede versioner $\tilde{X}_m$ af tids-frekvens-segmenterne X_m ;

9. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge krav 8, hvor den første og anden normaliserings- og/eller transformationsenhed er konfigureret til at anvende en eller flere af de følgende algoritmer til tids-frekvens-segmenterne henholdsvis X_m og S_m , almindeligvis betegnet som Z_m , hvor underindeks, tidsindeks m springes over for enkeltheds skyld i det følgende udtryk:

- normalisering af rækker til nul-middelværdi:

$$g_1(Z) = Z - \mu_z^r \mathbf{1}_z^T,$$

hvor

$$\mu_z^r$$

er en $J \times 1$ vektor, hvis j^{te} registrering er middelværdien af den j^{te} række af Z , deraf det øvre indeks r i

$$\mu_z^r,$$

, hvor $\mathbf{1}$ betegner en $N \times 1$ vektor af enere, og hvor øvre indeks T betegner matrixtransposition;

• normalisering af rækker til enhedsnorm:

$$\text{hvor } g_2(Z) = D^r(Z)Z,$$

$$D^r(Z) = \text{diag}\left(\left[1/\sqrt{Z(1,:)Z(1,:)^H} \quad \dots \quad 1/\sqrt{Z(J,:)Z(J,:)^H}\right]\right),$$

hvor $\text{diag}(\cdot)$ er en diagonalmatrix med elementerne af argumenterne på hoveddiagonalen, og hvor $Z(j,:)$ betegner den j^{te} række af Z , således at $D^r(Z)$ er en $J \times J$ diagonalmatrix med den omvendte norm af hver række på hoveddiagonalen, og nuller andetsteds, det øvre indeks H betegner Hermitian transposition, og for-multiplikation med $D^r(Z)$ normaliserer rækkerne af den resulterende matrix til enhedsnorm;

• Fourier transformation anvendes på hver række

$$g_3(Z) = ZF,$$

hvor F er en $N \times N$ Fourier-matrix;

• Fourier transformation anvendes på hver række efterfulgt af beregning af størrelsen af de resulterende kompleks-værdifulde elementer

$$g_4 = |ZF|$$

hvor $|\cdot|$ beregner de elementvise størrelser;

• identitetsoperatøren

$$g_5(Z) = Z.$$

• normalisering af kolonner til nul-middelværdi:

$$h_1(Z) = Z - \underline{\mathbf{1}}\mu_z^{c^T},$$

hvor

$$\mu_z^f$$

er en $N \times 1$ vektor, hvis i^{te} registrering er middelværdien af den i^{te} række af Z , og hvor $\underline{\mathbf{1}}$ betegner en $J \times 1$ vektor af enere;

• normalisering af kolonner til enhedsnorm:

$$\text{hvor } h_2(Z) = ZD^c(Z),$$

$$D^c(Z) = \text{diag}\left(\left[1/\sqrt{Z(:,1)^H Z(:,1)} \quad \dots \quad 1/\sqrt{Z(:,N)^H Z(:,N)}\right]\right),$$

hvor $Z(:,n)$ betegner den n^{te} række af Z , således at $D^c(Z)$ er en diagonal- $N \times N$ matrix med den omvendte norm af hver kolonne på hoveddiagonalen, og nuller andetsteds, og hvor en efter-multiplikation med $D^c(Z)$ normaliserer rækkerne af den resulterende matrix til enhedsnorm.

10. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge et hvilket som helst af kravene 1-9, hvor den mellemliggende taleforståelighedsberegningssenhed (ISIU) er indrettet til at bestemme de mellemliggende taleforståelighedscoefficients d_m i afhængighed af en

prøvekorrelationskoefficient $d(a,b)$ af elementerne i to $K \times 1$ vektorer a og b , hvor $d(a,b)$ defineres ved:

$$d(a,b) = \frac{\sum_{k=1}^K (a(k) - \mu_a)(b(k) - \mu_b)}{\sqrt{\sum_{k=1}^K (a(k) - \mu_a)^2 (b(k) - \mu_b)^2}},$$

hvor

$$\mu_a = \frac{1}{K} \sum_{k=1}^K a(k)$$

og

$$\mu_b = \frac{1}{K} \sum_{k=1}^K b(k),$$

hvor k er indekset for vektorregistreringen, og K er vektordimensionen, og hvor a og b repræsenterer K elementer fra tidsfrekvenssegmenter henholdsvis S_m eller $\tilde{S}_m$, og X_m eller $\tilde{X}_m$.

11. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge krav 10, hvor det mellemliggende forståelighedsindeks d_m defineres som

- den gennemsnitlige prøvekorrelationskoefficient af alle kolonner i henholdsvis S_m og X_m , eller $\tilde{S}_m$ og $\tilde{X}_m$, det vil sige,

$$d_m = \frac{1}{N} \sum_{n=1}^N d(\tilde{S}_m(:,n), \tilde{X}_m(:,n)),$$

n er et kolonneindeks, eller som

- den gennemsnitlige prøvekorrelationskoefficient af alle rækker i S_m og X_m , eller $\tilde{S}_m$ og $\tilde{X}_m$, det vil sige,

$$d_m = \frac{1}{J} \sum_{j=1}^J d(\tilde{S}_m(j,:), \tilde{X}_m(j,:))$$

j er et rækkeindeks, eller som

- prøvekorrelationskoefficienten af alle elementer i S_m og X_m , eller $\tilde{S}_m$ og $\tilde{X}_m$, det vil sige,

$$d_m = d(\tilde{S}_m(:,), \tilde{X}_m(:,)),$$

hvor notationen $S_m(:,)$ og $X_m(:,)$, eller $\tilde{S}_m(:,)$ og $\tilde{X}_m(:,)$, repræsenterer $NJ \times 1$ vektorer, der er dannet ved stabling af kolonnerne af de respektive matricer.

12. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge et hvilket som helst af kravene 1-11, hvor den endelige taleforståelighedsberegningssenhed (FSIU) er indrettet til at beregne den endelige taleforståelighedsforudsigelse d fra de mellemliggende taleforståeligheds-koefficienter d_m , eventuelt transformeret af en funktion $u(d_m)$, som et gennemsnit over tid af informationssignalet x :

$$d = \frac{1}{M} \sum_{m=1}^M u(d_m)$$

hvor M repræsenterer varigheden i tidsenheder af den taleaktive del af den støjende og/eller behandlede version x af målsignalet.

13. Monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge krav 12, hvor funktion $u(d_m)$ defineres som

$$u(d_m) = \log\left(\frac{1}{1-d_m^2}\right),$$

5 eller som

$$u(d_m) = d_m.$$

10 14. Høreapparat (HD), der er indrettet til at blive placeret ved eller i venstre eller højre øre af en bruger, eller til at blive helt eller delvist implanteret i hovedet af brugeren, hvor høreapparatet omfatter en monaural taleforståelighedsforudsigelsesenhed (MSIP) ifølge et hvilket som helst af kravene 1-13.

15 15. Høreapparat (HD) ifølge krav 14, der er konfigureret til adaptivt at modificere behandlingen af et indgangssignal (y) til høreapparatet (HD) for at maksimere den monaurale taleforståelighedsforudsigelse d , for at forbedre brugerens forståelighed af et udgangssignal (u) af høreapparatet (HD), der præsenteres for brugeren (U).

20 16. Fremgangsmåde til tilvejebringelse af en monaural taleforståelighedsforudsigelse for at estimere en brugers evne til at forstå et informationssignal x , der omfatter en støjende og/eller behandlet version af et måltalsignal, hvor fremgangsmåden omfatter

- 25 • tilvejebringelse af en tids-frekvens-repræsentation $s(k,m)$ for den støjfrie version s af målsignalet, hvor k er et frekvensområdeindeks, $k=1, 2, \dots, K$, og m er et tidsindeks;
- tilvejebringelse af en tids-frekvens-repræsentation $x(k,m)$ af den støjende og/eller behandlede version x af målsignalet, hvor k er et frekvensområdeindeks, $k=1, 2, \dots, K$, og m er et tidsindeks;
- 30 • tilvejebringelse af en tids-frekvens-subbånds-repræsentation $s_j(m)$ for den støjfrie version s af målsignalet, der repræsenterer temporale indhyllinger, eller funktioner deraf, af frekvens-subbånds-signaler $s_j(m)$ for det støjfrie målsignal, j er et frekvens-subbånds-indeks, $j=1, 2, \dots, J$, og m er tidsindekset;
- 35 • tilvejebringelse af en tids-frekvens-subbånds-repræsentation $x_j(m)$ for den støjende og/eller behandlede version x af målsignalet, der repræsenterer temporale indhyllinger, eller funktioner deraf, af frekvens-subbånds-signaler $x_j(m)$ af den støjende og/eller behandlede version af målsignalet, $j=1, 2, \dots, J$, og m er tidsindekset;
- 40 • deling af tids-frekvens-subbånds-repræsentationen $s_j(m)$ for den støjfrie version s af målsignalet til tids-frekvens-segmenter S_m , der svarer til et antal N af på hinanden følgende prøver af subbånds-signalerne;
- 45 • deling af tids-frekvens-subbånds-repræsentationen $x_j(m)$ for den støjende og/eller behandlede version x af målsignalet til tids-frekvens-segmenter X_m , der svarer til et antal N af på hinanden følgende prøver af subbånds-signalerne;

hvor den første og anden tids-frekvens-segmentdelingsenhed er konfigureret til henholdsvis at dele tids-frekvens-repræsentationerne $s_j(m)$ og $x_j(m)$ i segmenter i form af spektrogrammer, der svarer til N på hinanden følgende prøver af alle subbånds-signaler, hvor det m^{te} segment er defineret af $J \times N$ matrixen

$$Z_m = \begin{bmatrix} z_1(m-N+1) & \cdots & z_1(m) \\ \vdots & & \vdots \\ z_J(m-N+1) & \cdots & z_J(m) \end{bmatrix},$$

hvor z , Z repræsenterer henholdsvis s , S eller x , X ;

- tilvejebringelse af i det mindste en normaliserings- og/eller transformationsfunktion af rækker og i det mindste en normaliserings- og/eller transformationsfunktion af kolonner af tids-frekvens-segmenterne S_m og X_m ;

- tilvejebringelse af mellemliggende taleforståelighedskoefficienter d_m , der estimerer en forståelighed af tids-frekvens-segmentet X_m , hvor de mellemliggende taleforståelighedskoefficienter d_m er baseret på de i det væsentlige støjfrie, normaliserede og/eller transformerede tids-frekvens-segmenter $\tilde{S}_m$ og de støjende og/eller behandlede, normaliserede og/eller transformerede tids-frekvens-segmenter $\tilde{X}_m$;

- beregning af en endelig taleforståelighedsforudsigelse d , der estimerer en forståelighed af den støjende og/eller behandlede version x af målsignalet ved at kombinere, for eksempel ved gennemsnit eller at anvende en MIN eller MAX-funktion, de mellemliggende taleforståelighedskoefficienter d_m , eller en transformeret version deraf, over tid.

DRAWINGS

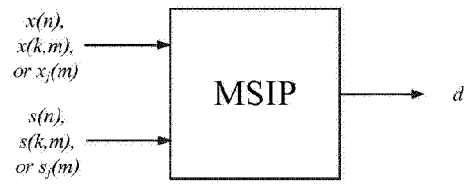


FIG. 2A

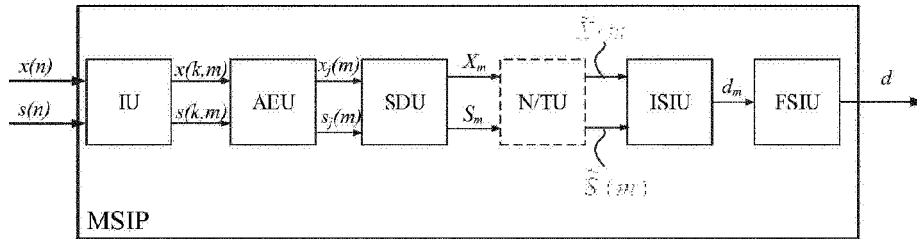


FIG. 2B

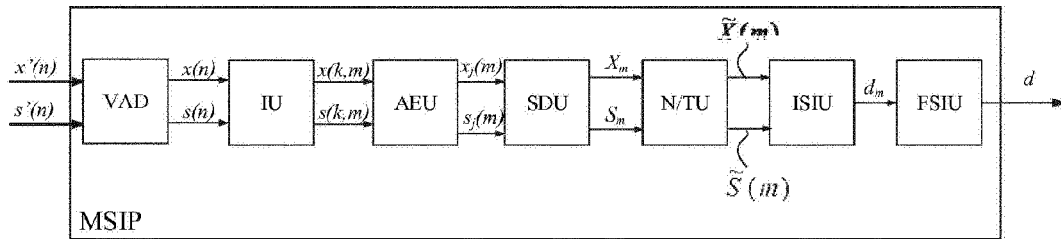


FIG. 2C

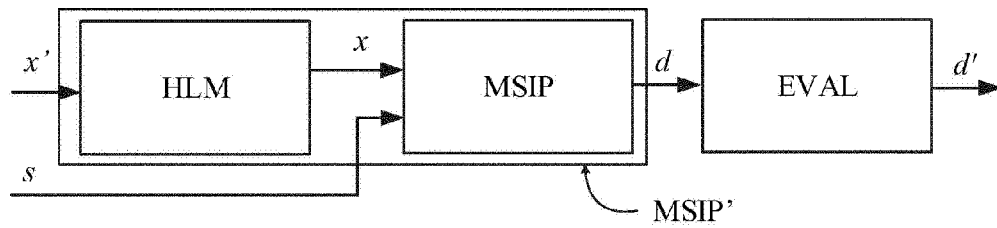


FIG. 3A

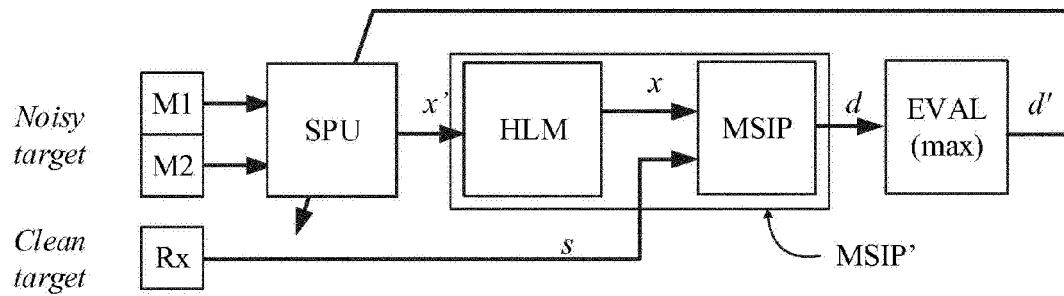


FIG. 3B

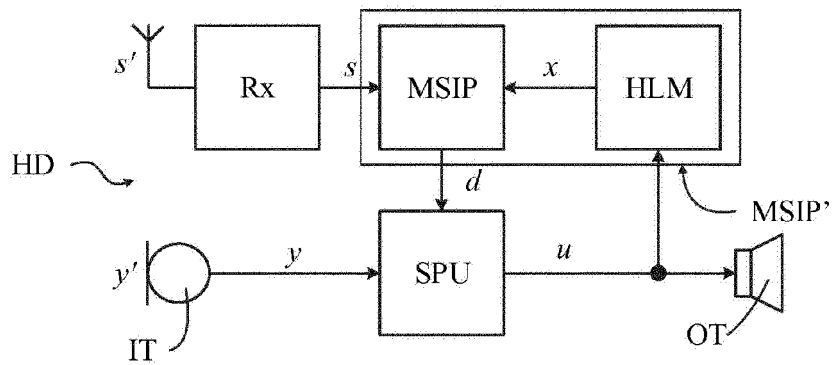


FIG. 3C

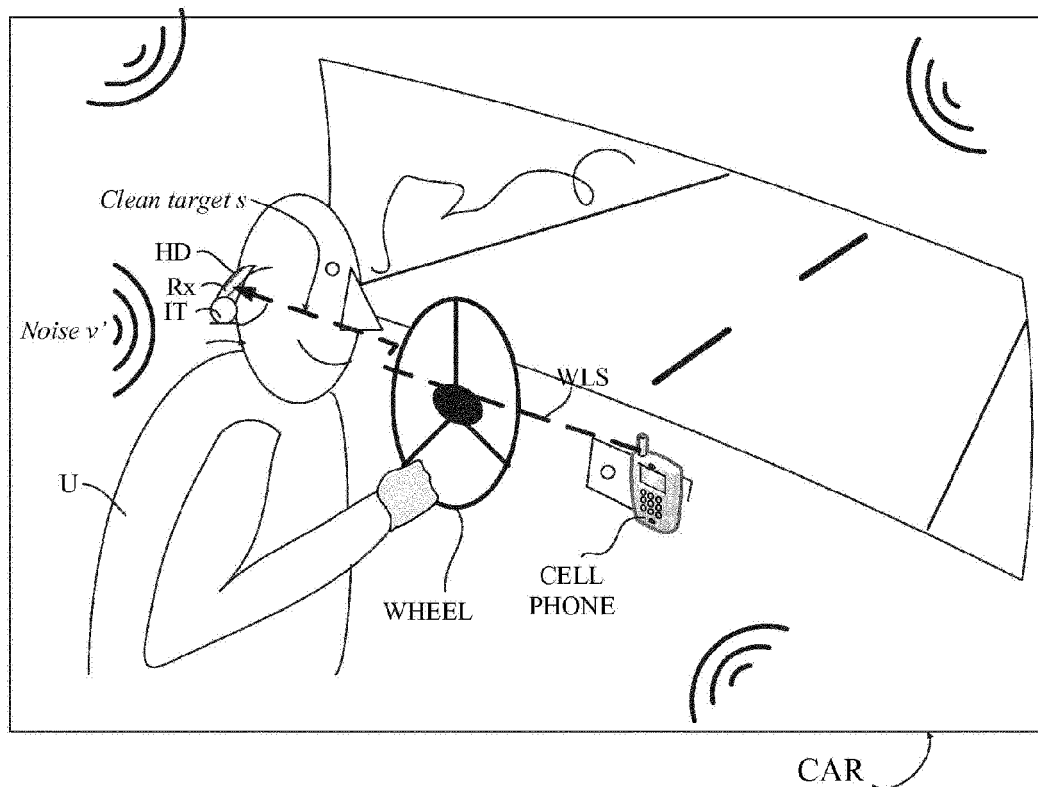


FIG. 5A

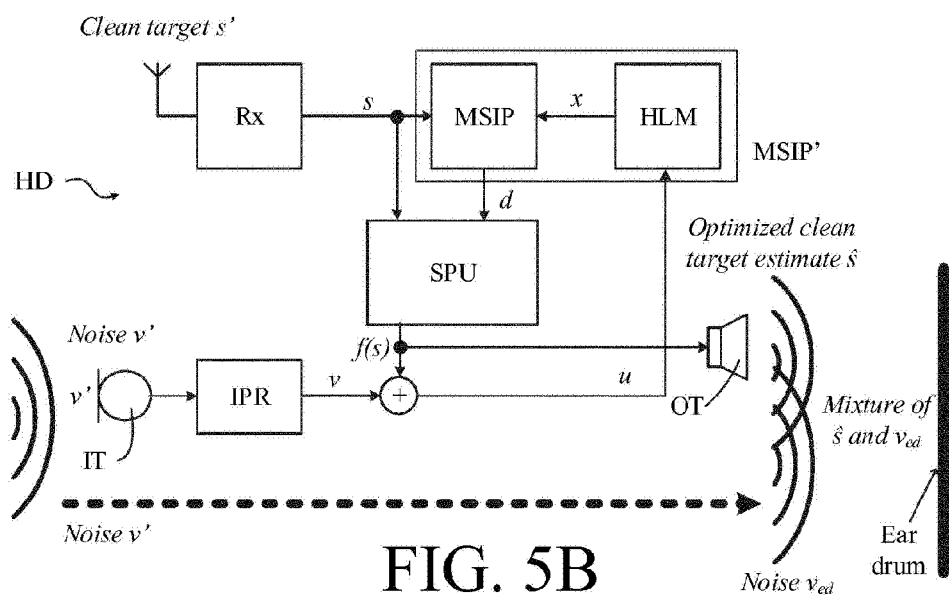


FIG. 5B

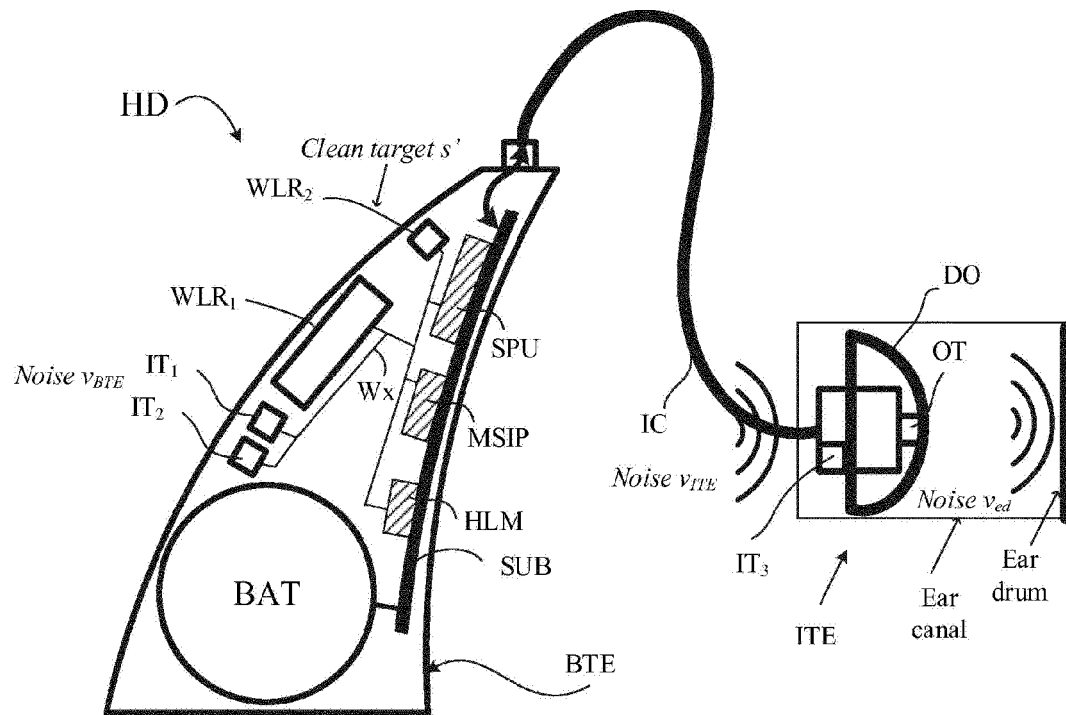


FIG. 5C

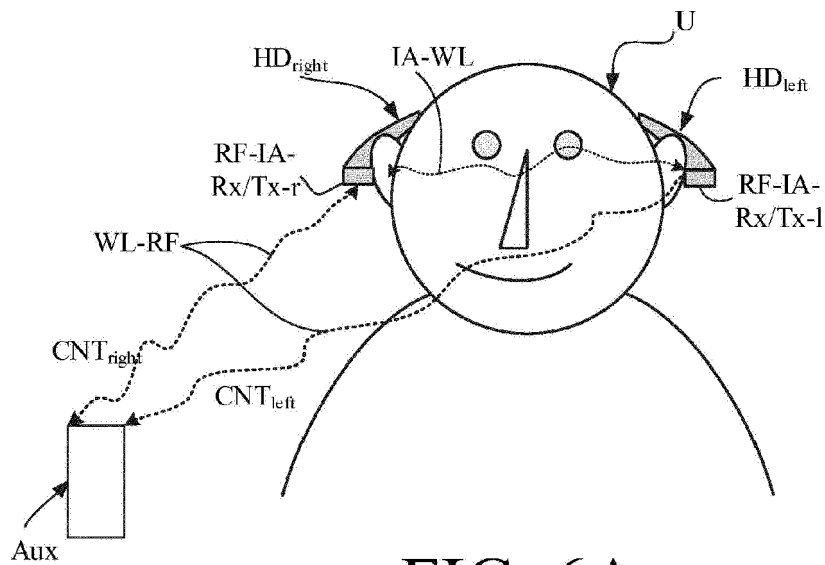


FIG. 6A

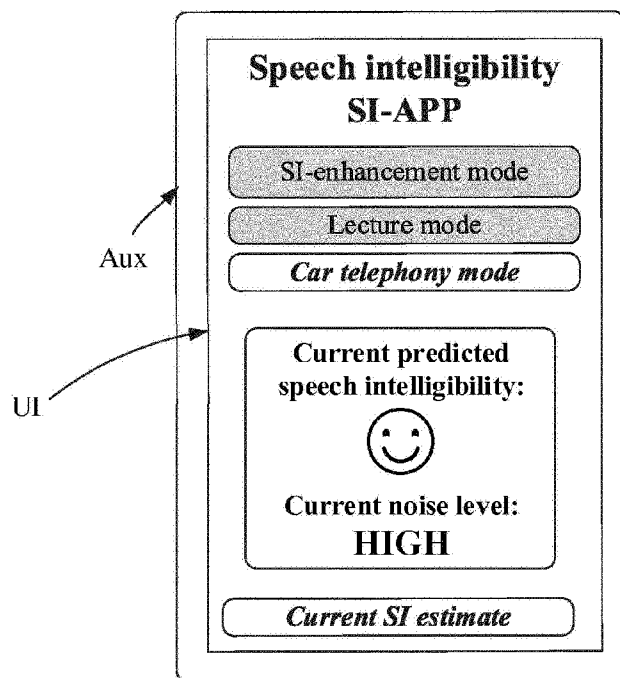


FIG. 6B

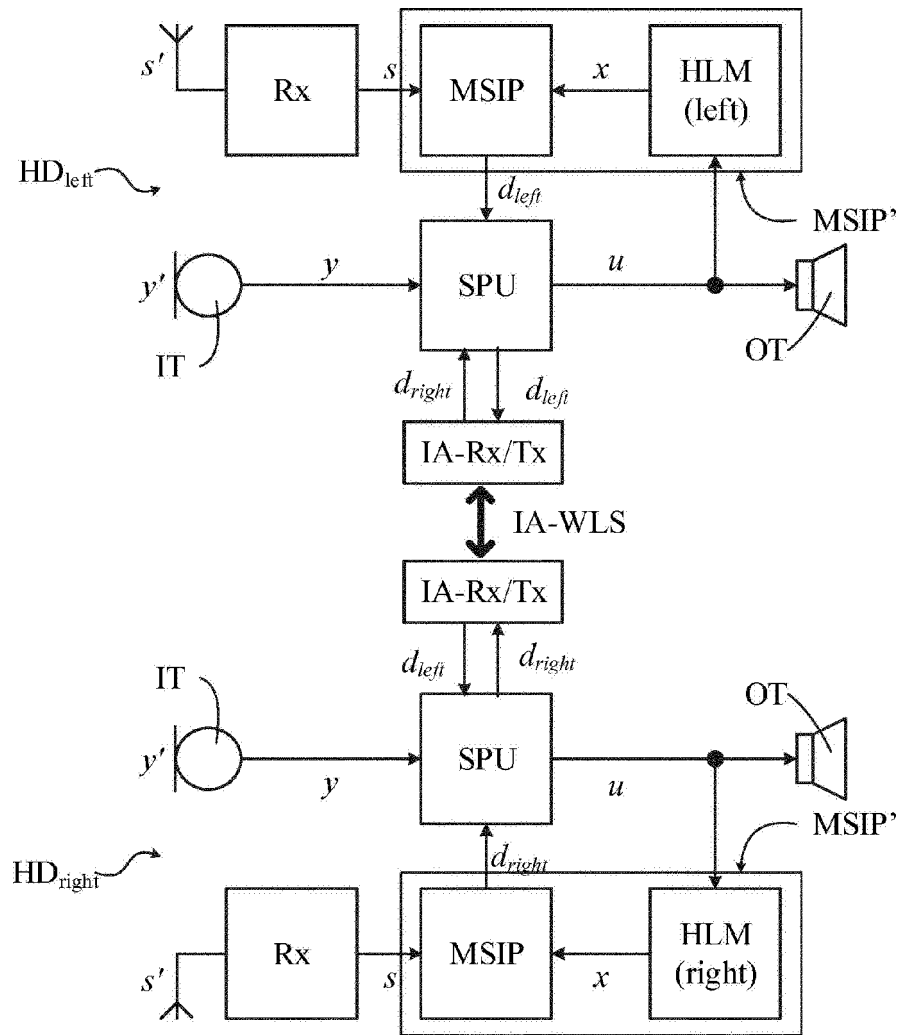


FIG. 7