OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(12) (19) (CA) **Demande-Application**

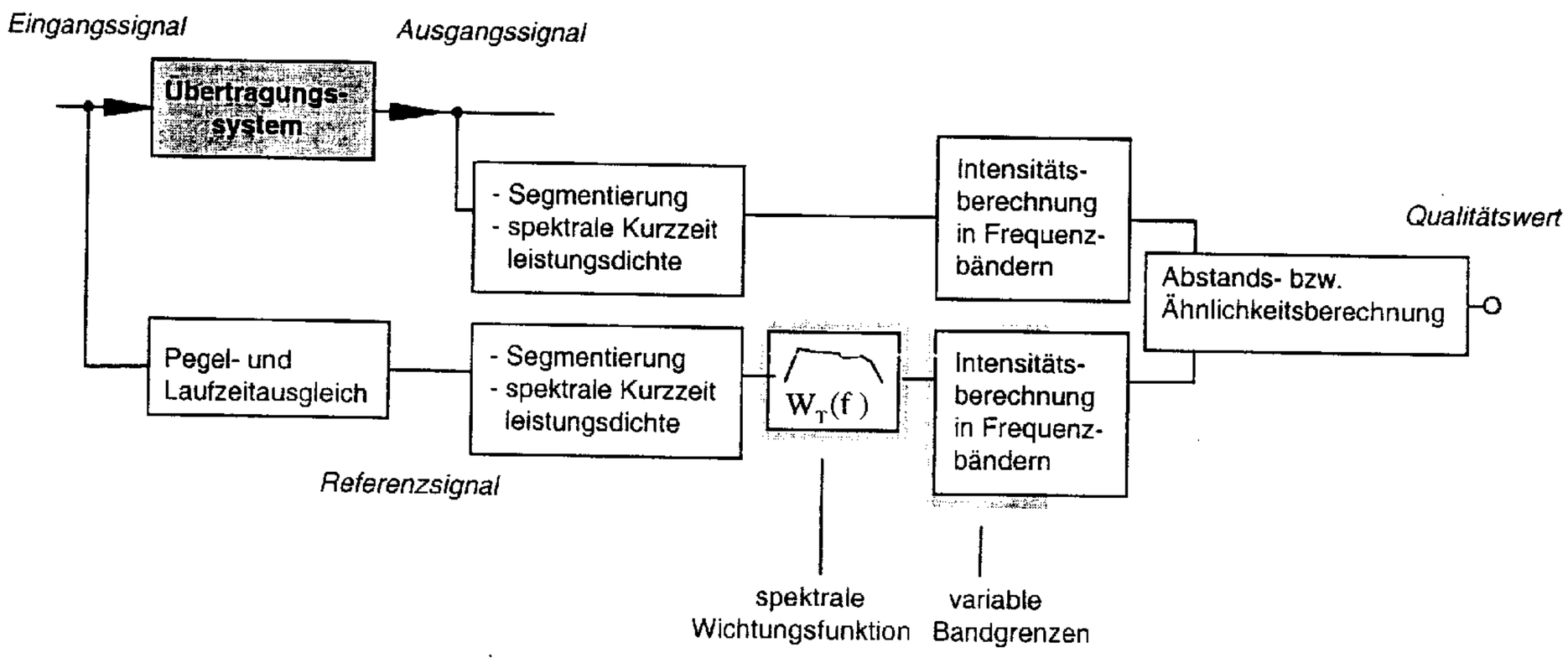
(21) (A1) **2,305,652**
(86) 1999/08/14
(87) 2000/03/09

(72) BERGER, JENS, DE
(71) DEUTSCHE TELEKOM AG, DE

(51) Int.Cl.[7] G10L 19/00
(30) 1998/08/27 (198 40 548.0) DE

(54) **PROCEDE DE DETERMINATION INSTRUMENTALE DE LA QUALITE VOCALE**

(54) **METHOD FOR INSTRUMENTAL VOICE QUALITY EVALUATION**



Eingangssignal = Input signal

Ausgangssignal = Output signal

Übertragungssystem = Transmission system

Pegel- und... = Level and transit time equalization

Referenzsignal = Reference signal

Segmentierung = Segmentation

spektrale... = Spectral short-time power density

spektrale Wichtungs... = Spectral weighting function

Intensitäts... = Intensity calculation in frequency bands

variable... = Variable band limits

Qualitätswert = Quality value

Abstands... = Calculation of distance/similarity

OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(21) (A1) **2,305,652**
(86) 1999/08/14
(87) 2000/03/09

(57) Les procédés connus de détermination instrumentale de la qualité vocale fondés sur une comparaison d'intensités du signal vocal à évaluer avec un signal vocal de référence n'évaluent pas de manière optimale des déformations spectrales du signal vocal à évaluer, de sorte que la détermination de la qualité n'est pas fiable. De plus, l'intégration de l'intensité du signal dans des bandes de fréquences à limites de bande constantes entraîne des évaluations erronées de biaisements déterminés du signal vocal à évaluer, comme des systèmes de codage à débits binaires réduits en induisent. Afin d'améliorer la fiabilité de proposition des paramètres de qualité calculés, on corrige d'une part dans une large mesure avec une fonction de pondération $W_T(f)$, des déformations de l'enveloppe spectrale moyenne du signal avant de procéder à une comparaison des propriétés spectrales. Les limites fixes de bande sont par ailleurs augmentées, afin d'intégrer la puissance volumique spectrale, et on recherche, pour les remplacer, dans une plage d'optimisation prédéfinie, des limites de bandes où les représentations d'intensité spectrales obtenues du signal vocal à évaluer et du signal de référence présentent une similitude maximale. Les solutions décrites permettent d'élargir des procédés connus et peuvent être ajoutées à leur structure.

(57) Known methods for instrumental voice quality evaluation based on comparing signal intensities of the voice signal to be evaluated with a reference voice signal do not optimally evaluate spectral distortions in the voice signal to be evaluated so that quality evaluation is unreliable. Moreover, by integrating the signal intensity in the frequency bands with constant band limits, certain falsifications of the voice signal to be evaluated, such as those caused, for instance, by coding systems with lower bit rates, are erroneously evaluated. In order to enhance prediction reliability of the evaluated quality parameters, distortions of the mean spectral envelope are extensively corrected with a weighting function $W_T(f)$ before comparing spectral properties. On the other hand, the fixed band limits for integration of spectral power density are suppressed and other band limits are searched for instead in a predetermined optimization area in which the resulting spectral intensity representations of the voice signal to be evaluated and the reference voice signal have maximum similarity. The solutions described can supplement known methods and can be incorporated into their structures.

Abstract

1. Process for instrumental speech-quality determination

2.1 Known processes for instrumental speech-quality determination based on a comparison of signal intensities of the speech signal to be assessed with a reference speech signal do not optimally assess spectral deformations of the speech signal to be assessed, with the result that the quality assessment is unreliable. Furthermore, as a result of the integration of the signal intensity in frequency bands with constant band limits, defined distortions of the speech signal to be assessed, caused, for example, by coding systems with low bit rates, are incorrectly assessed.

2.2 In order to improve the reliability of the calculated quality characteristic values, first, deformations of the mean spectral envelopes are extensively corrected with a weighting function $W_T(f)$ prior to comparison of the spectral properties. Second, the fixed band limits for integration of the spectral power density are removed and, instead, within a given optimization range, band limits are found at which the resulting spectral intensity representations of speech signal to be assessed and reference speech signal exhibit maximum similarity.

2.3 The described solutions may expand known processes and may be added to the structure thereof.

3. Fig. 2a

Process for instrumental speech-quality determination

Description

Preliminary remarks

The invention relates to a process for instrumental ("objective") speech-quality determination in which characteristic values for speech-quality determination are derived by a comparison of properties of a speech signal to be assessed with properties of a reference speech signal (undisturbed signal).

Usually, the speech-quality determination of speech signals is carried out by means of auditory ("subjective") examinations with test persons.

The aim of instrumental ("objective") processes for speech-quality determination is, using suitable calculation methods, to use properties of the speech signal which is to be assessed to establish characteristic values which describe the speech quality of the speech signal to be assessed without having to resort to the judgments of test persons.

The calculated characteristic values and the underlying process for instrumental speech-quality determination are regarded as acknowledged if there is a high correlation with the results of auditory reference examinations. Consequently, the speech-quality values obtained by means of auditory examinations represent the target values which are to be achieved by instrumental processes.

Prior art

Known processes for instrumental speech-quality determination are based on the comparison of a reference speech signal with the speech signal which is to be assessed, the reference speech signal and the speech signal to be assessed being segmented into short time portions. In said segments, the spectral properties of the two signals are compared. Various approaches and models are used to calculate the spectral short-time properties. Usually, the signal intensity is calculated in frequency bands, the width of which becomes greater with increasing mid-frequency. Examples of such frequency bands are the known

third-octave bands or critical bands according to Zwicker (published in Zwicker, E.: "Psychoakustik" ["Psychoacoustics"], Berlin: Springer publishing house, 1982).

The thus calculated spectral intensity representation for each time portion under consideration can be viewed as a series of numerical values in which the number of individual values corresponds to the number of frequency bands used, the numerical values themselves representing the calculated intensity values and a consecutive index of the frequency bands describing the sequence of the numerical values.

In the presently known processes for instrumental speech-quality determination, the limits of the frequency bands used are kept constant on the frequency axis.

In each time segment under consideration, the calculated intensities of speech signal to be assessed and of reference speech signal are compared with each other in each band. The difference of both values, or the similarity of the two resulting spectral intensity representations, constitutes the basis for the calculation of a quality value (Fig. 1).

Such processes have been developed in particular for the qualitative assessment of speech in telephony applications. Examples thereof are the publications:

"A perceptual speech-quality measure based on a psychacoustic sound representation" (Beerends, J. G.; Stemerdink, J. A., J. Audio Eng. Soc. 42(1994)3, pp.115-123)

"Auditory distortion measure for speech coding" (Wang, S; Sekey, A.; Gersho, A.: IEEE Proc. Int. Conf. acoust., speech and signalprocessing (1991), pp.493-496).

The presently valid ITU-T standard P.861 likewise describes such a process: "Objective quality measurement of telephone-band speech codecs" (ITU-T Rec. P.861, Geneva 1996).

Disadvantages of known instrumental speech-quality measurement processes

The use of known processes for instrumental speech-quality determination falls down on the reliability of the calculated quality values for defined signal properties which are to be assessed. Presently known processes deliver unreliable quality values especially in the case

of impairments in the speech signal which is to be assessed, for example in the case of impairments caused by speech coding processes with low bit rates or combinations of different disturbances.

In such cases, the presently known processes have the disadvantage that, in the case of a comparison between the speech signal to be assessed and a reference speech signal, the quality characteristic value which is to be calculated includes differences between the two signal portions in the selected representation plane which either do not lead or scarcely lead to a qualitative impairment, not even one which is perceptible in the auditory test.

Within the framework of the herein considered transmission of speech in telephone applications, frequency-band limitations and spectral deformations of the speech signal to be assessed (caused, for example, by filter properties of the telephone device or of the transmission channel) make only a limited contribution to a perceived qualitative impairment.

In order partially to prevent such deficiencies, an attempt is made in a different approach to compensate for the linear distortions (frequency response) by a correction filter or power transmission function (published in: "A new approach to objective quality-measures based on attribute-matching", Halka, U.; Heute, U., Speech communication, 11(1992)1, pp.15-30). However, the use of said process is disadvantageous in the case of nonlinear and time-invariant transmission, because the thus calculated compensation function no longer describes exclusively the spectral deformations of the signal which is to be assessed. Displacements of spectral short-time maxima ("formant displacements") in the signal under test in relation to the reference speech signal, caused, for example, by coding systems with low bit rates, lead in known processes to large differences in the spectral intensity representations and therefore have a great impact on the calculated quality value. Investigations have revealed that, in an auditory speech-quality examination, however, such displacements of spectral short-time maxima have only a limited influence on the quality verdict.

Problem

The object of the invention is to reduce the influence of spectral limitations and deformations of the speech signal to be assessed and also the influence of displacements of spectral short-time maxima prior to the comparison of the spectral properties of a signal to be tested with a reference speech signal and prior to the calculation of a quality value in instrumental processes.

Solution

In contrast to known approaches, generated in the herein described invention is a spectral weighting function which is based on mean spectral envelopes, e.g. the mean spectral power density, of speech signal to be assessed and reference speech signal. This permits the use of the process likewise in the case of nonlinear and time-variant transmission.

The spectral weighting function is calculated from the quotients of the given values of the mean spectral power density of the signal to be assessed $Phi_Y(f)$ and of that of the input signal of the transmission system $Phi_X(f)$ such that the weighting function can be described via

$$\mathbf{W_T(f) = a(f) \cdot (Phi_Y(f) / Phi_X(f))}.$$

The assessment function a(f) can weight the weighting function $W_T(f)$ differently over the range of effect, being, in the simplest case, constant at 1.

The thus calculated spectral weighting function $W_T(f)$ approximates the mean spectral envelopes of speech signal to be assessed and reference speech signal, with the result that differences of the two spectral envelopes are included only to a reduced extent in the calculated quality value.

The spectral weighting function $W_T(f)$ can be applied, firstly, to the reference speech signal, the reference speech signal being approximated in its mean spectral power density to the signal to be assessed (Fig. 2a).

Secondly, the spectral weighting function can be applied, inverted, to the signal to be assessed. The distortion of the latter is thereby eliminated and, with regard to its mean spectral power density, it is approximated to the reference speech signal (Fig. 2b).

A further part of the invention relates to the correction of displacements of spectral short-time maxima which are caused by the transmission systems.

The intensity is integrated for each time portion in frequency bands. The result is a series of intensity values for each spectral representation of a signal portion, each individual value representing the intensity in a frequency band. In this connection, the displacements of spectral short-time maxima may lead to different calculated intensities in the frequency bands of reference speech signal and speech signal to be assessed.

Said differences in the spectral intensity representations - caused by displacements of spectral short-time maxima - can be reduced by a variable arrangement of the frequency bands on the frequency axis. In contrast to the constant band limits in known processes, the band limits are displaced on the frequency axis. However, the number of frequency bands and the index thereof remain constant. In an optimization loop, those band limits are then accepted at which the two resulting spectral representations of speech signal to be assessed and reference speech signal exhibit maximum similarity or whose distance from each other is the smallest. Such optimization is carried out for all bands in all time segments under consideration.

The use of variable band limits to calculate the spectral intensity representation is not restricted simply to the signal in which the described spectral weighting function $W_T(f)$ is also used, but may also be applied to the other signal and even to both signals (see Fig. 2a and 2b).

Example embodiment:

A special example embodiment is shown by a realization according to Fig. 3, which is known as TOSQA (Telecommunication Objective Speech Quality Assessment), in which there is extended preprocessing of the reference speech signal.

In specifications of the general realizations according to Fig. 2a and 2b, speech pauses are detected by means of a speech pause detector and are not included in the quality measure. Likewise, the reference speech signal and the speech signal to be assessed are filtered with a 300...3400 Hz bandpass filter and there is also filtering to the frequency response of a telephone handset. The integration of the spectral power density is carried out in critical bands which represent the basis for the calculation of the specific loudness.

However, the integration in critical bands is *not* in fixed critical band limits, but with the variable critical band limits described in the present invention. The calculated signal powers in the thus modified critical bands form the basis for intensity calculation. Use was made in this respect of a model for calculating the specific loudness according to Zwicker, an aurally compensated intensity representation (published in Zwicker, E.: "Psychoakustik" ["Psychoacoustics"], Berlin: Springer publishing house, 1982).

As an addition to the general approach, the calculated loudness patterns are supplemented by an error assessment function. The calculated quality value is formed via a mean value of the correlation coefficients of the specific loudnesses for each short time segment under consideration over the number of evaluated speech segments.

What is claimed is: (6)

1. Process for instrumental speech-quality determination in which characteristic values for speech-quality determination are calculated by a comparison of spectral short-time properties of a speech signal to be assessed with a reference speech signal,

characterized in that

prior to comparison of the properties of the speech signals, differences in mean spectral envelopes are reduced in that calculated initially therefrom is a spectral weighting function with which are weighted the spectral short-time properties of the speech signals in all time segments under consideration, with the result that the differences in the mean spectral envelopes are thereby included only to a limited extent in the quality characteristic value which is to be calculated; and

for the calculation of the signal intensity, the limits of the frequency bands used are made variable, with the result that, for each signal portion under consideration in all evaluated frequency bands, the calculated intensities of reference speech signal and signal to be assessed differ as little as possible from each other.

2. Process according to claim 1, characterized in that, first, the mean spectral envelopes of speech signal to be assessed and reference speech signal are calculated in the form of a mean power density spectrum and a spectral weighting function $W_T(f)$ is calculated from the quotients of both spectra, the short-time power density spectra of the reference speech signal being weighted with said spectral weighting function $W_T(f)$ prior to the calculation of a quality characteristic value.

3. Process according to claims 1 and 2, characterized in that the weighting function $W_T(f)$ to be calculated is calculated only from partial regions of the calculated mean spectral envelopes of speech signal to be assessed and reference speech signal and, consequently, the differences in mean spectral envelopes between both signals are reduced only in spectral partial regions.

4. Process according to claims 1 to 3, characterized in that, prior to calculation of the quality characteristic values, there is an integration of the signal intensity for each evaluated short time portion in critical bands, the limits of the critical bands being

variable on the frequency axis, but the width of the critical bands remaining constant on the pitch scale, wherein the specific loudness is calculated from the signal intensities in the critical bands, the limits of those critical bands being used in which the calculated differences in specific loudness between the signal to be assessed and the reference speech signal have the smallest difference in the band and time segment under consideration.

5. Process according to claims 1 to 4, characterized in that the quality characteristic value is calculated from the similarity of the spectral representations in each time portion under consideration, the similarity representing a correlation coefficient, averaged over all time portions under consideration, between the spectral representation of the speech signal to be assessed and the spectral representation of the reference speech signal in the respective time segment.

6. Process according to claim 5, characterized in that the correlation coefficient between the spectral representation of the speech signal to be assessed and the spectral representation of the reference speech signal in the respective time segment is calculated from only a partial region of the spectral representation, i.e. not all calculated spectral values are taken into consideration for the calculation of the quality characteristic value.

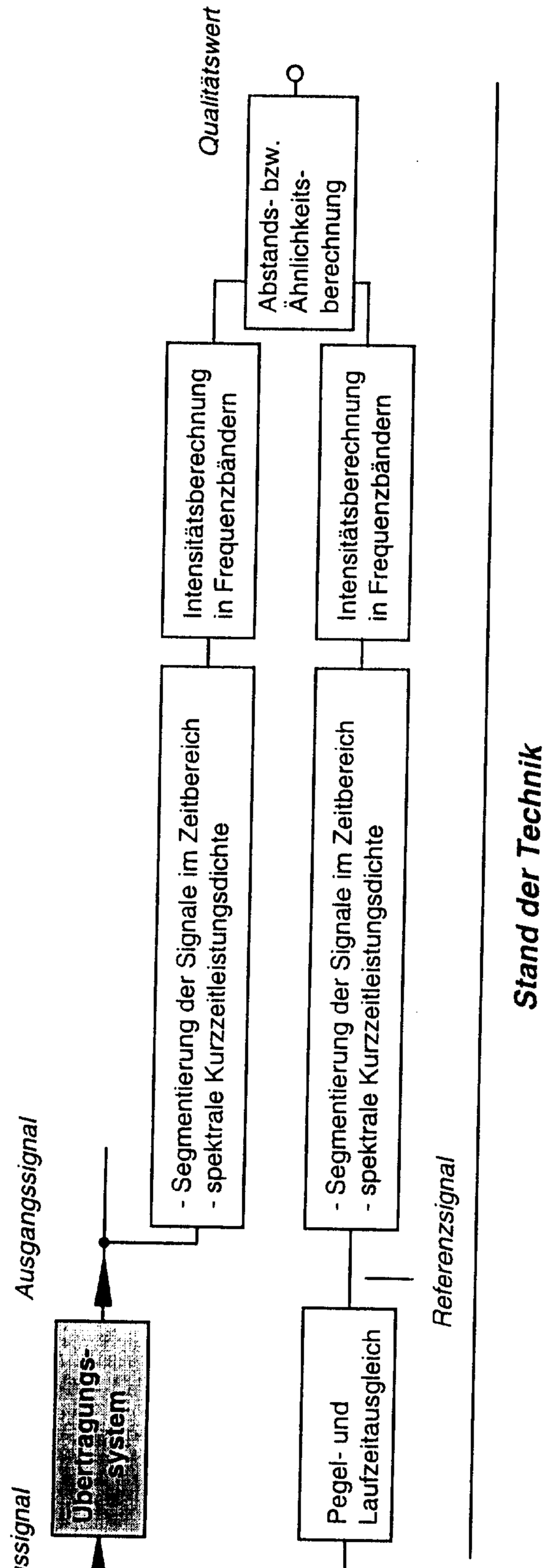

Qualitätswert
Abstands- bzw. Ähnlichkeits- berechnung
Intensitätsberechnung in Frequenzbändern
Intensitätsberechnung in Frequenzbändern
- Segmentierung der Signale im Zeitbereich
- spektrale Kurzzeitleistungsdichte
- Segmentierung der Signale im Zeitbereich
- spektrale Kurzzeitleistungsdichte
Ausgangssignal
Eingangssignal
Übertragungs- system
Pegel- und Laufzeitausgleich
Referenzsignal
Stand der Technik

Fig. 1

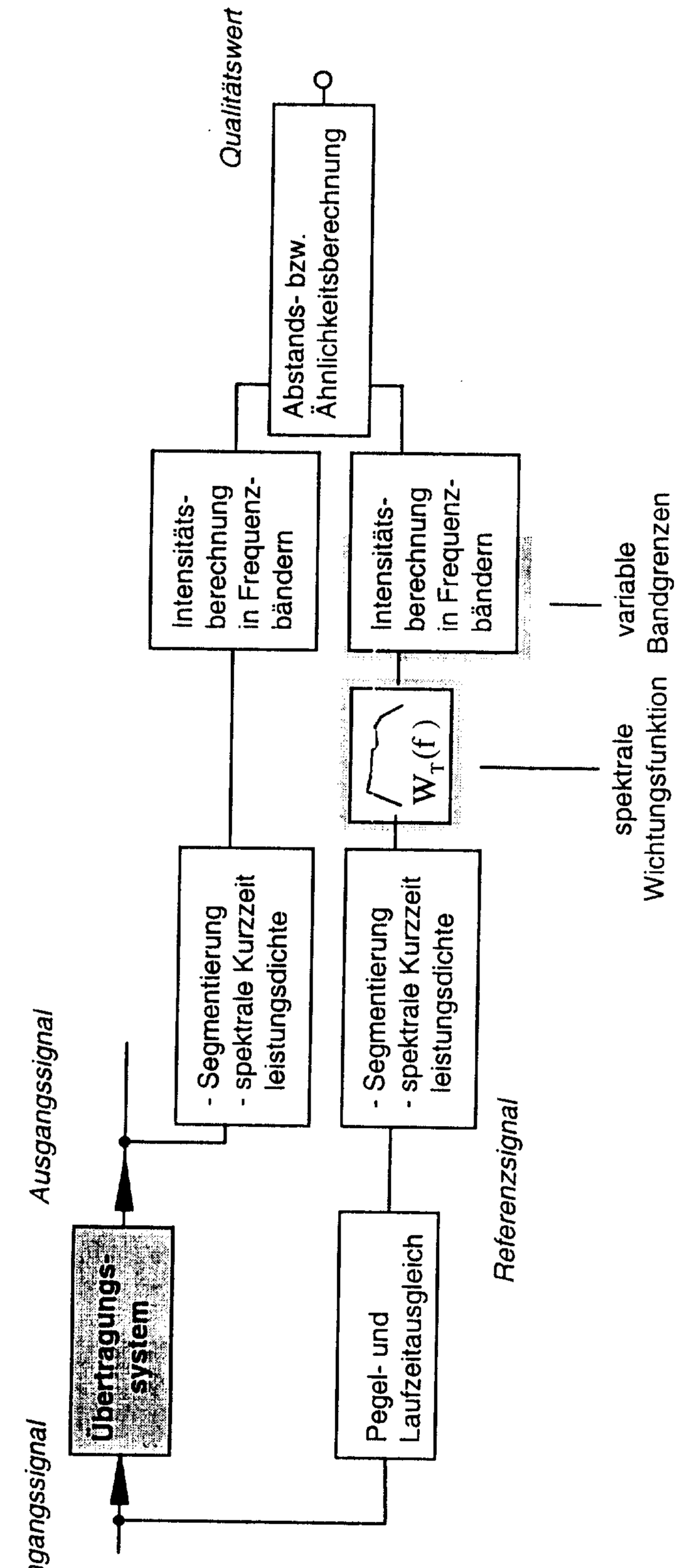

Qualitätswert
Abstands- bzw. Ähnlichkeitsberechnung
Intensitäts-berechnung in Frequenz-bändern
Intensitäts-berechnung in Frequenz-bändern
$W_T(f)$
spektrale
variable Bandgrenzen
Wichtungsfunktion
- Segmentierung
- spektrale Kurzzeit leistungsdichte
- Segmentierung
- spektrale Kurzzeit leistungsdichte
Pegel- und Laufzeitausgleich
Referenzsignal
Ausgangssignal
Übertragungs-system
Eingangssignal

Fig. 2a

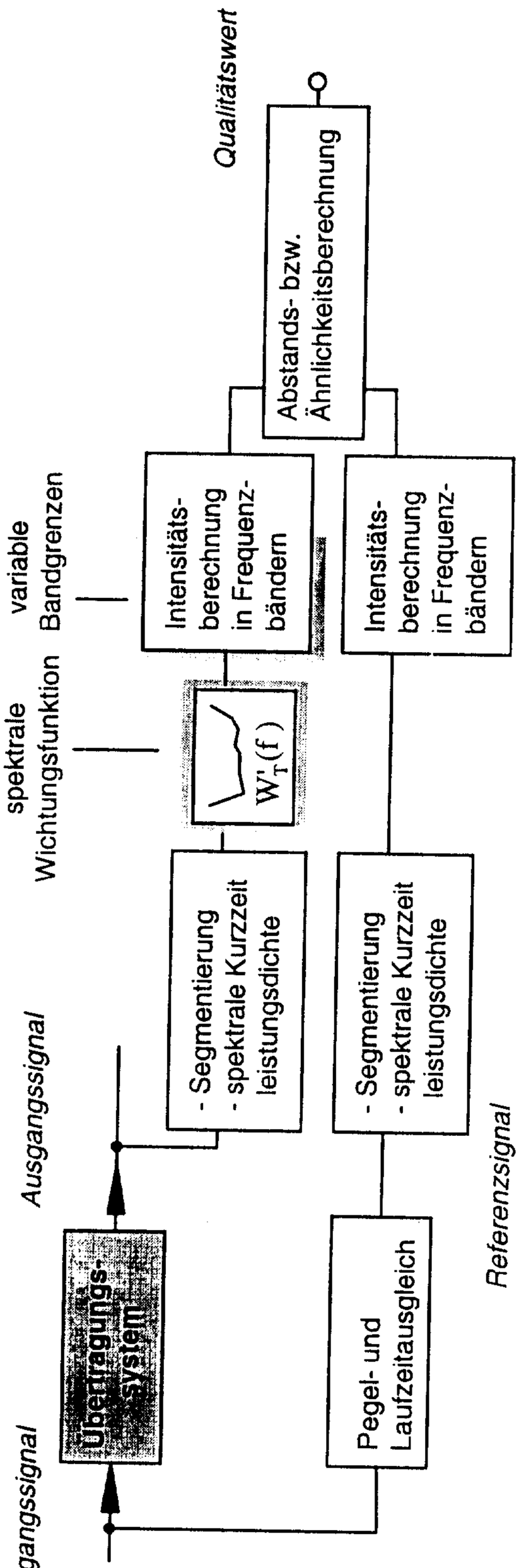

Qualitätswert
Abstands- bzw. Ähnlichkeitsberechnung
variable Bandgrenzen
Intensitäts-berechnung in Frequenz-bändern
Intensitäts-berechnung in Frequenz-bändern
spektrale Wichtungsfunktion
$W'_T(f)$
- Segmentierung
- spektrale Kurzzeit leistungsdichte
- Segmentierung
- spektrale Kurzzeit leistungsdichte
Ausgangssignal
Referenzsignal
Übertragungs-system
Pegel- und Laufzeitausgleich
Eingangssignal

Fig. 2b

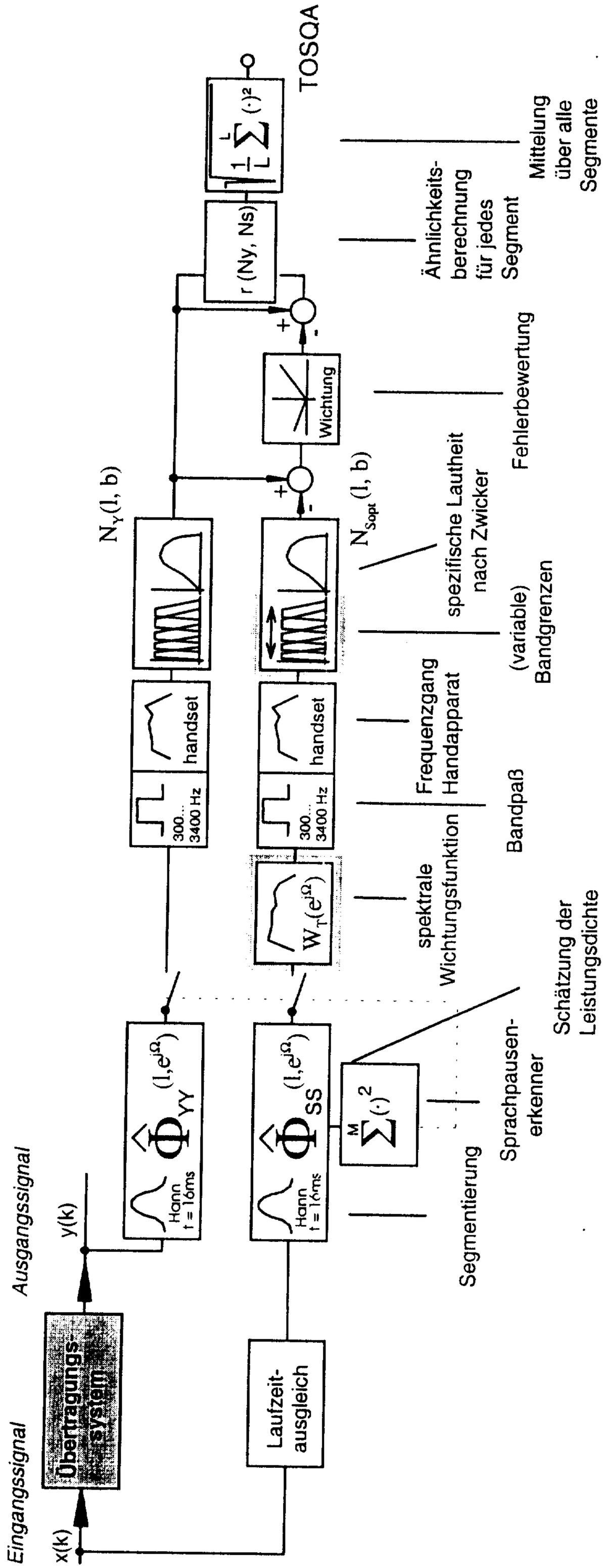

TOSQA
$\frac{1}{L}\sum_{L}(\cdot)^2$
r (Ny, Ns)
Ähnlichkeits-
berechnung
für jedes
Segment
Mittelung
über alle
Segmente
$N_Y(l, b)$
handset
300...
3400 Hz
Wichtung
$N_{Sopt}(l, b)$
Frequenzgang
Handapparat
spezifische Lautheit
nach Zwicker
Fehlerbewertung
Bandpaß
(variable)
Bandgrenzen
$W_T(e^{j\Omega})$
spektrale
Wichtungsfunktion
$\hat{\Phi}_{YY}(l,e^{j\Omega})$
$\hat{\Phi}_{SS}(l,e^{j\Omega})$
Hann
t = 16ms
$\sum_{M}(\cdot)^2$
Segmentierung
Sprachpausen-
erkenner
Schätzung der
Leistungsdichte
Ausgangssignal
y(k)
Übertragungs-
system
Eingangssignal
x(k)
Laufzeit-
ausgleich

Fig. 3

Legenden zu den Abbildungen:

Fig. 1

Eingangssignal = Input signal

Ausgangssignal = Output signal

Übertragungssystem = Transmission system

Pegel- und... = Level and transit time equalization

Referenzsignal = Reference signal

Segmentierung... = Segmentation of signals in time

spektrale... = Spectral short-time power density

Intensitäts... = Intensity calculation in frequency bands

Qualitätswert = Quality value

Abstands... = Calculation of distance/similarity

Stand der Technik = Prior art

Fig. 2a

Eingangssignal = Input signal

Ausgangssignal = Output signal

Übertragungssystem = Transmission system

Pegel- und... = Level and transit time equalization

Referenzsignal = Reference signal

Segmentierung = Segmentation

spektrale... = Spectral short-time power density

spektrale Wichtungs... = Spectral weighting function

Intensitäts... = Intensity calculation in frequency bands

variable... = Variable band limits

Qualitätswert = Quality value

Abstands... = Calculation of distance/similarity

Fig, 2b:

Eingangssignal = Input signal

Ausgangssignal = Output signal

Übertragungssystem = Transmission system

Pegel- und... = Level and transit time equalization

Referenzsignal = Reference signal

Segmentierung = Segmentation

spektrale... = Spectral short-time power density

spektrale Wichtungs... = Spectral weighting function

Intensitäts... = Intensity calculation in frequency bands

variable... = Variable band limits

Qualitätswert = Quality value

Abstands... = Calculation of distance/similarity

Fig. 3:

Eingangssignal = Input signal

Ausgangssignal = Output signal

Übertragungssystem = Transmission system

Laufzeitausgleich = Transit time equalization

Segmentierung = Segmentation

Sprachpausenerkenner = Speech pause detector

Schätzung... = Estimation of power density

spektrale Wichtungs... = Spectral weighting function

Bandpaß = Bandpass

Frequenzgang... = Frequency response of handset

(variable)... = (Variable) band limits

spezifische Lautheit... = Specific loudness according to Zwicker

Wichtung = Weighting

Fehlerbewertung = Error assessment

Aehnlichkeits... = Calculation of similarity for each segment

Mittelung... = Averaging over all segments