



(19)

REPUBLIK
ÖSTERREICH
Patentamt

(10) Nummer: AT 407 584 B

(12)

PATENTSCHRIFT

(21) Anmeldenummer: 1084/99 (51) Int. Cl.⁷: G10L 15/00
(22) Anmeldetag: 21.06.1999 G10L 17/00, 13/02, 19/02, 19/04,
(42) Beginn der Patentdauer: 15.08.2000 21/00
(45) Ausgabetag: 25.04.2001

(56) Entgegenhaltungen:
US 4389540A US 4860359A US 5150413A

(73) Patentinhaber:
BERNHARD HANS-PETER DIPLO.ING. DR.TECHN.
A-4300 ST. VALENTIN, OBERÖSTERREICH (AT).

(72) Erfinder:
BERNHARD HANS-PETER DIPLO.ING. DR.TECHN.
ST. VALENTIN, OBERÖSTERREICH (AT).

(54) VERFAHREN ZUR DYNAMISCHEN UND STATISCHEN DARSTELLUNG VON SPRACHSIGNALEN

AT 407 584 B

(57) Die dynamische und statische Darstellung von Sprachsignalen erfolgt durch eine zwei oder mehrdimensionale Visualisierung auf zwei oder mehrdimensionalen Anzeigegeräten, wie zum Beispiel einem Bildschirm. Auf den Koordinaten der Anzeigegeräte werden das Originalsignal und gleichzeitig das um eine bestimmte Zeitverzögerung verschobene Signal angelegt. Die so entstehenden Bilder können bereits in der zweidimensionalen Darstellung phonetisch interpretiert werden. Die Resultate können für das Analysieren von Sprachsignalen oder auch zum visuellen Lernen des Sprechens herangezogen werden. Die Zeitverzögerung wird bei diesem Verfahren mit Hilfe des ersten Minimums der Transformation zwischen zwei verschiedenen Signalen bestimmt.

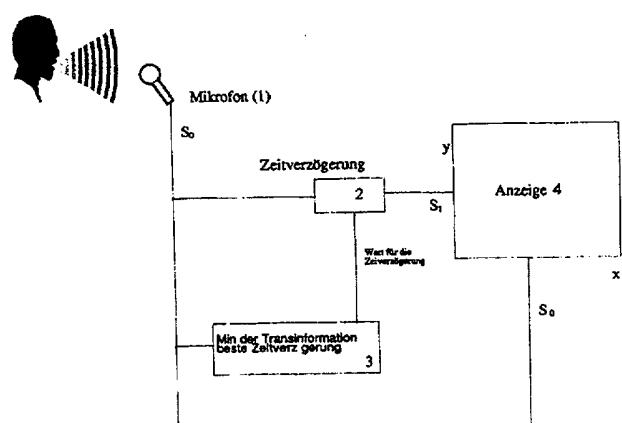


Fig. 1: Verfahren

Gegenstand der Erfindung ist ein Verfahren zur dynamischen und statischen Darstellung von Sprachsignalen, bestehend aus einer geeigneten Zusammenschaltung von Zeitverzögerungen und einem Anzeigegerät.

Es sind verschiedene Verfahren zur Darstellung von Sprachsignalen bekannt: Zeitverhalten der Amplitude, Zeitverhalten des Frequenzspektrums (Spektrogramm) und einige mehr. Auf diese Verfahren wird nicht näher eingegangen, da sie als allgemein bekannt vorausgesetzt werden können.

Dem Erfinder ist das Patent US 4389540 A (Nakamura et al.) bekannt, und die in Fig. 3-6 gezeigten und im zugehörigen Text beschriebenen Zusammenhänge beziehen sich auf die Anpassung von digitalen Filtern die zur Bestimmung der linearen Prädiktionsfilterkoeffizienten dienen. Dieser Vorgang bezieht sich ausschließlich auf die optimale Bestimmung der Filterkoeffizienten und nicht auf die Visualisierung des Signales.

Dem Erfinder ist das Patent US 4860359 A (Eicher) bekannt, und die in Fig. 1-3 gezeigten und im zugehörigen Text beschriebenen Zusammenhänge beziehen sich auf das sprachgesteuerte Übertragungssystem, das insbesondere dazu dient Sprachübertragung zu kontrollieren und nicht dazu um Sprache zu visualisieren, da kein wie immer geartetes Anzeigemedium für das Sprachsignal vorgesehen ist. Die erwähnte Zeitverzögerung bezieht sich auf den Start der Übertragung.

Dem Erfinder ist das Patent US 5150413 A (Nakamura et. al.) bekannt, und die in Fig. 5-6 gezeigten und in der Zusammenfassung beschriebenen Zusammenhänge beziehen sich ausschließlich auf lineare Filter und nicht auf die Anzeige der Signale. Die in der Zusammenfassung erwähnten phonetischen Eigenschaften sind spektrale Eigenschaften, die in der gegenständlichen Erfindung keine Rolle spielen.

Die gegenständliche Erfindung vermeidet, im Gegensatz zu den bekannten Verfahren, die fehlende Beschreibung des zugrunde liegenden dynamischen Systems. Die Darstellung ermöglicht eine phonetische Interpretation der entstehenden Graphen, die auf das erzeugende System zurückgeführt werden können.

Der Erfindung liegt die Aufgabe zugrunde, Sprachsignale so aufzubereiten, dass diese von einem Beobachter oder Meßsystem zum Zwecke der Untersuchung oder Nachbildung des laubildenden Vorgangs herangezogen werden können. In einer konkreten Anwendung ist vorstellbar, dass sich die Darstellung als Sprachlerngerät für hörbehinderte Menschen eignen kann. In einer weiteren Anwendung ist es möglich, dass die so aufgezeichneten Graphen die Basis für lernende Algorithmen bilden, die das Sprachsignal als dynamisches System modellieren können.

Vorteilhaft gegenüber allen anderen bekannten Verfahren ist hier, dass die Dynamik der Sprachlaute in einem 2 oder 3-dimensionalen Bild beschrieben werden kann. Diese Abbildung geht auf die dynamische Systemtheorie zurück. Dynamische Systeme werden mittels Differentialgleichungen beschrieben. Diese Beschreibung fehlt aber bei Sprachsignalen, da wir den Erzeugungsmechanismus nicht als Gleichung gegeben haben. Es ist daher notwendig, aus den bekannten Daten (Sprachsignal) die Dynamik des Systems zu rekonstruieren. Im Fall der Differentialgleichungen wird immer von Ableitungen des Signals ausgegangen. Da im gegenständlichen Fall eine Messung vorliegt, wird durch additives Messrauschen die Methode der höheren Ableitungen (überproportionales Verstärken von hochfrequenten Störsignalanteilen) für die Signalrekonstruktion unbrauchbar. Es wird daher die Zeitverzögerungsmethode verwendet die von Takens für allgemeine dynamische Systeme eingeführt wurde. Es sind aber für Sprachsignale keine Werte der Zeitverzögerung bekannt. Für die Zeitverzögerung wird jener Wert verwendet, bei dem die Information zwischen dem Originalsignal und dem verzögerten Signal am geringsten ist. Für die weiteren Koordinaten wird ein Vielfaches der zuerst gefundenen Verzögerung verwendet. Zum Bestimmen der gegenseitigen Information wird der Mutual Information angewendet, den der Autor in schon in *Elektrotechnik und Informationstechnik*, 111(12):648-649, Dez. 1994, präsentiert hat.

Nach einem vorteilhaften Merkmal der Erfindung ist vorgesehen, daß die dynamische Abfolge von gesprochenen Lauten in einem Bild dargestellt werden kann, da der entstandene Graph in einer bestimmten Anzeigeskala (pro Koordinate) die Kurven darstellt. Die Zeit ist entlang der Kurven aufgetragen und nicht entlang einer Koordinate, die sich in eine Richtung ausdehnt. So ist zum Beispiel für die Darstellung eines Lautüberganges mittels spektraler Darstellung das Signal in kurze Signalabschnitte zu zerlegen (windowing), um diese spektral zu analysieren und dann hintereinander anzusehen. Bei der gegenständlichen Erfindung wird diese Segmentierung ver-

mieden und in einer kompakten Darstellung kann der gesamte phonetische Prozess des Laut-überganges dargestellt werden.

Ein Ausführungsbeispiel des Verfahrens ist in den Zeichnungen und der folgenden Beschreibung dargestellt. Es zeigt:

5 Die Sprachsignale werden mittels eines Mikrofons aufgenommen und über ein, zwei oder mehrere Zeitverzögerungsglieder (Zeitverzögerung τ_1, τ_2, \dots) geführt. Das Originalsignal soll in weiterer Folge S_0 heißen und die verzögerten Signale S_1, S_2, \dots . Die Zeitverzögerung wird mit Hilfe des in Fig.1 gezeigten Verfahrens eingestellt. In Fig.1 wird ein Mikrofon (1) gezeigt, dass das Originalsignal S_0 aufnimmt. Dieses Signal wird in der beispielhaften Skizze für 2 - dimensionale 10 Darstellung der x-Koordinate der Anzeige (5) zugeführt. Die y-Koordinate der Anzeige (4) wird durch ein Signal S_1 versorgt, das über die Zeitverzögerung (2) geführt wurde. Die Zeitverzögerung wird durch Minimieren der Transinformation in Block (3) bestimmt.

15

PATENTANSPRÜCHE:

1. Verfahren zur dynamischen Visualisierung von Sprachsignalen in zwei - oder drei - dimensionalen Darstellungen, dadurch gekennzeichnet, dass das elektrisch aufgenommene Sprachsignal, zum Beispiel mittels eines Mikrofons (1), ein oder zweimal zeitverzögert (2) wird und das Originalsignal S_0 auf der x Achse und das zugehörige verzögerte Signal auf der y Achse des Anzeigemediums (4), zum Beispiel Computerbildschirm oder Oszilloskop, aufgetragen ist und wenn eine dritte Koordinate des Anzeigemediums vorhanden ist, diese mit dem Signal S_2 versorgt wird, also mit dem um τ_2 verzögerten Signal.
2. Verfahren zur Wahl der Zeitverzögerung für das Visualisierungsverfahren nach Anspruch 1 dadurch gekennzeichnet, dass jene Zeitverzögerung gewählt wird, bei der die Unabhängigkeit beider Signale maximal ist, wobei dieses Maximum durch Bestimmen der Transinformation zwischen den beiden Signalen ermittelt wird und die Signale so lange gegeneinander zeitlich verschoben werden, bis die Transinformation ein erstes Minimum zeigt (3) und diese Zeitverzögerung sodann als τ_1 verwendet wird.

30

HIEZU 1 BLATT ZEICHNUNGEN

35

40

45

50

55

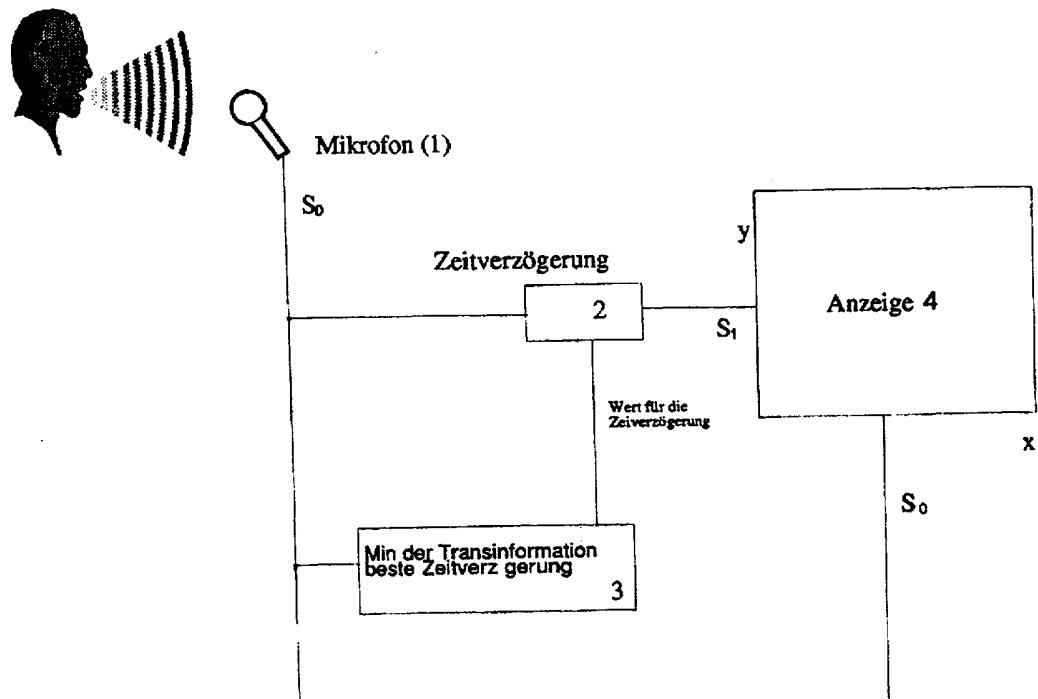


Fig. 1: Verfahren