

# Akustische Charakteristiken der Normalstimme

Jannis Kiosses und Hans Werner Strube  
Drittes Physikalisches Institut, Universität Göttingen

## Einleitung

Bei der Untersuchung von Stimmpathologien und der Klassifikation von gestörten Stimmen wird als Referenzgruppe die ungestörte Stimme (Normalstimme) herangezogen, die sich dadurch auszeichnet, dass sie ohne medizinischen Befund ist und akustisch normal klingt. Im Allg. ist diese Normalstimme jedoch nicht näher definiert. Es wird nur festgehalten, dass sie nicht pathologisch ist. Andererseits wird die interessierende Pathologie genauer charakterisiert und es wird festgehalten, worin sie sich von der gewählten Referenzgruppe (oder -stimme) unterscheidet, ohne jedoch diese Referenzgruppe (also die Normalstimme) entsprechend akkurat zu bestimmen (vergl. [1]).

Dagegen soll an dieser Stelle ein Satz von Messgrößen zusammengestellt werden, der die Normalstimme bestimmt. Somit können gestörte Stimmen besser abgegrenzt werden und eventuell kann man sogar die spezielle Störung in gewissen Grenzen bestimmen.

## Material

Aus einer Stimmdatenbank von rund 2000 Stimmaufnahmen wurde eine zufällige Auswahl von pathologischen Stimmen (76 Probanden und Pathologien) herausgegriffen und diese einer Anzahl (32 Probanden) attestierter Normalstimmen gegenüber gestellt.

Jede Aufnahme umfasst zunächst drei Vokalreihen zu je sieben isoliert gesprochenen Vokalen [ $\varepsilon$ ;, a;, e;, i;, o;, u;,  $\varepsilon$ ] in gehaltener Phonation. Die erste Vokalreihe wurde in normaler, die zweite in tieferer und die dritte in höherer Stimmlage gesprochen. Darauf folgte eine vierte Vokalreihe wieder in normaler Stimmlage, jedoch nach Stimmbelastung durch das Vorlesen eines Textes („Nordwind und Sonne“) von ca. 2 Minuten Dauer.

Die Stimmaufnahmen erfolgten digital (Pioneer D-07 DAT-Rekorder) in schallisolierter Umgebung. Die Analyse der eingeschwungenen Stimmsignale erfolgte automatisch jeweils auf 500ms-Intervallen mit einem Vorschub von jeweils 250ms.

## Methode

Für jedes Signal-Intervall aus den Aufnahmen wurden insgesamt sechs Grundeigenschaften bestimmt.

Durch Anwendung der Autokorrelation auf das jeweilige Stimmsignal-Intervall und Mittelung über das gesamte Signal wurde die mittlere Periodenlänge abgeschätzt. Schwankungen im Zeitsignal werden durch Kreuzkorrelation von aufeinander folgenden Signal-Intervallen bestimmt (mean waveform matching coefficient, MWC). Der MWC ist somit ein Maß für die Stabilität der Grundfrequenz im Stimmsignal.

Zur Ermittlung des Jitters und des Shimmers wird aus dem Signalstück die mittlere Periodenlänge (und ihre Schwankung) bzw. die mittlere Periodenamplitude (und ihre Schwankung) jeweils über  $n$  Perioden bestimmt (mit  $n=2; 3; 5; 7; 11; 15$ ). So erhält man eine Aussage sowohl über die kurzfristigen Schwankungen der Signalperiode bzw. Signalamplitude ( $n$  sei klein; für wenige Perioden) als auch über die langfristige Tendenz der Signalschwankungen ( $n$  sei groß; für viele Perioden).

Die Rauschgrößen NNE (normalized noise energy) und CHNR (cepstral harmonic to noise ratio) werden jeweils für drei unterschiedliche Frequenzbänder (60-5000Hz; 60-2000Hz; 1000-5000Hz) bestimmt. Der GNE-Wert (glottal to noise excitation ratio; ein Maß für den Glottisschluss) wird für drei Bandbreiten (1000Hz; 2000Hz; 3000Hz) bestimmt. Der GNE beruht auf der Korrelation der Hilbertenüllenden des inversgefilterten Signalstücks zwischen verschiedenen Frequenzkanälen.

Zur näheren Erläuterung der Messgrößen und ihrer Dimensionen, siehe auch [1],[2]

Die so ermittelten insgesamt 22 Messgrößen je Signal-Intervall werden als Datenvektor in einem 22-dimensionalen Punktraum interpretiert (siehe auch [1]). Durch Anwendung von Kohonen-Karten (einer speziellen Klasse von künstlichen neuronalen Netzen) wird eine allgemeine Klassifikation der Messdaten durchgeführt. Dabei wird eine begrenzte Anzahl von Kohonen-Karten-Vektoren durch einen Abstandsminimierung iterativ an die Datenvektorverteilung im Punktraum angepasst.

Mittels der Fischerschen Diskriminanzanalyse werden alle Messgrößen nach Relevanz für die Gruppentrennung sortiert. Es wird auch ein Trennmaß bestimmt, das eine Aussage über die Signifikanz der Gruppentrennung durch die verwendeten Messgrößen liefert.

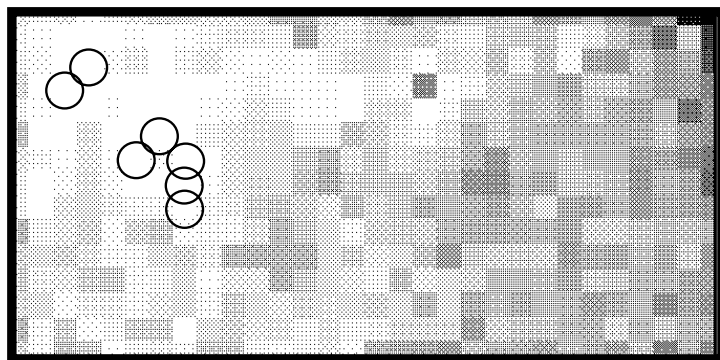


Abbildung 1: Stimmlokalisierungen auf der Kohonen-Karte. Die eingekreisten Zellen erkennen nur Normalstimmen.

## Ergebnisse

Die Klassifizierung der Stimmgruppen (Normalstimmen, Pathologien) durch die Kohonen-Karten liefert eine grobe Zuordnung der Messgrößen zu den zwei Stimmgruppen. In Abb.1 ist beispielhaft eine 15x30-Zellen Karte dargestellt. Jede Karten-Zelle repräsentiert eine spezielle Parameterkonstellation (bzw. Pathologie). Die eingekreisten Zellen der Karte sind ausschließlich für Normalstimmen empfindlich (Normalzellen). Wie auch in Abb.1 zu erkennen ist, werden die Normalzellen in zwei Untergruppen unterteilt. Zum einen die obere Untergruppe, hier mit zwei Normalzellen und zum anderen die untere Untergruppe die sich aus fünf Normalzellen formiert.

Die die Normalzellen bestimmenden Parameterwerte werden statistisch ausgewertet. Das Ergebnis ist ein allgemeiner Parametersatz der die zwei Untergruppen der Normalstimme charakterisiert (siehe dazu Tab.1.).

Die Fischersche Diskriminanzanalyse wiederum liefert eine nach Signifikanz für die Gruppentrennung sortierte Liste (s. Tab.2) aller Messgrößen. Für die Tab.2 wurde jedoch die mittlere Signifikanz über beide Untergruppen der Normalstimme verwendet, da die jeweiligen Messgrößen-Signifikanzen der Untergruppen sich nur unwesentlich unterscheiden.

Somit kann Anhand der Tab.2 eine Mindestanzahl von Messgrößen gewählt werden, um eine Normalstimme signifikant von einer speziellen Stimmstörung oder einer Auswahl von Stimmstörungen zu unterscheiden. In kritischen Fällen können weitere Messgrößen hinzugezogen werden, was einer Erhöhung der Auflösung entspräche. Die Gütetabelle (s. Tab.3) zeigt die erreichbare Trenngüte im Vergleich zur Anzahl der berücksichtigten Pa-

parameter. Die Auswahl des jeweils hinzuzunehmenden Parameters erfolgt nach Tab.2.

Maß	obere Gruppe.		untere Gruppe.	
	Mittel	Stdabw.	Mittel	Stdabw.
MWC	0.9968	0.0002	0.996	0.001
Jitter2	0.1282	0.0039	0.25	0.06
Jitter3	0.0558	0.0037	0.121	0.033
Jitter5	0.0769	0.0034	0.152	0.035
Jitter7	0.0944	0.0009	0.185	0.041
Jitter11	0.1274	0.0037	0.232	0.047
Jitter15	0.1574	0.0067	0.264	0.048
Shim.2	1.184	0.034	1.74	0.25
Shim.3	0.5263	0.0003	0.79	0.13
Shim.5	0.689	0.013	1.04	0.15
Shim.7	0.82	0.03	1.27	0.18
Shim.11	1.074	0.057	1.64	0.23
Shim.15	1.316	0.079	1.94	0.26
GNE1	0.9823	0.0047	0.9773	0.0048
GNE2	0.9692	0.0061	0.9577	0.0067
GNE3	0.9367	0.0002	0.9035	0.0089
NNE1	-26.56	0.52	-18.6	1.9
NNE2	-28.77	0.94	-21	2
NNE3	-28.7	1.9	-17.2	1.8
CHNR1	45.70	0.82	39.9	1.6
CHNR2	49.56	0.35	42.6	1.6
CHNR3	24.0	2.5	17.7	1.9

**Tabelle 1:** Aus der Kohonen-Karte ermittelte charakteristische Werte für die beiden Untergruppen der Normalstimme.

## Diskussion

Aus der Kohonen-Karte werden die Parameterwerte für die allgemeine Normalstimme (die Primärstimme) ermittelt (Tab.1). Hier ist zu beachten, dass Werte der Tab.1 jeweils bestimmend für die Primärstimme sind. D.h. wenn von den ausgewählten Parameter einer von den Tabellenwerten abweichen, handelt es sich beim Probanden wahrscheinlich nicht um eine Primärstimme.

Die zwei Untergruppen der Primärstimme sind vermutlich geschlechtsspezifisch, was aber nicht zwingend ist. Es sind auch andere Spezifikationen für die Untergruppen denkbar, wie z.B. das Alter oder Stimmtraining und so weiter. Hier ist eine weitergehende Untersuchung notwendig.

Die Gütetabelle (Tab.3) zeigt, dass schon die ersten Parameter ausreichen um die Primärstimme zu bestimmen. Die Kombination von mehr als acht Parameter hat hier keinen wesentlichen Einfluss auf die Trenngüte. Somit kann eine Begrenzung der Parameterzahl auf die notwendigsten für eine Stimmuntersuchung vorgenommen werden. Jedoch gilt im allgemeinen, dass bestimmte Pathologien der Normalstimme sehr ähnlich sein können. In dem Fall kann die notwendige Anzahl von Parametern die für die benötigte Trennschärfe gebraucht wird auch größer sein.

Rang	Maß	Rang	Maß
1	GNE3	12	GNE2
2	Shimmer3	13	Jitter5
3	NNE3	14	CHNR2
4	CHNR3	15	CHNR1
5	NNE1	16	NNE2
6	MWC	17	Shimmer7
7	Jitter2	18	Shimmer2
8	Jitter3	19	GNE1
9	Shimmer15	20	Shimmer11
10	Jitter7	21	Shimmer5
11	Jitter15	22	Jitter11

**Tabelle 2:** Durch die Diskriminanzanalyse erstellte Rangliste der signifikantesten Maße zur Normalstimmen-Bestimmung.

Anzahl Maße	Güte in %.	add. Maß
1	71.55	GNE3
2	74.33	Shimmer3
3	75.15	NNE3
4	75.81	CHNR3
5	76.31	NNE1
6	76.87	MWC
7	76.98	Jitter2
8	77.43	Jitter3
9	77.42	Shimmer15
10	77.42	Jitter7
11	77.56	Jitter15
12	77.41	GNE2
13	77.63	Jitter5
14	77.63	CHNR2
15	77.72	CHNR1
16	77.62	NNE2
17	77.72	Shimmer7
18	77.66	Shimmer2
19	77.64	GNE1
20	77.69	Shimmer11
21	77.65	Shimmer5
22	77.66	Jitter11

**Tabelle 3:** Verbesserung der Diskriminanzgüte durch Hinzunehmen eines weiteren Maßes nach der Rangliste.

## Zusammenfassung

Zur allgemeinen Charakterisierung der Gruppe der Normalstimmen wurden eine Reihe von Stimmaufnahmen (mit Normalstimmen und gestörten Stimmen) mittels Kohonen-Karten klassifiziert. Somit konnte eine allgemeine Unterscheidung zwischen den zwei Stimmgruppen Normalstimme und gestörten Stimmen erreicht werden.

Aus der Betrachtung derjenigen Kohonen-Zellen auf der Karte, die ausschließlich für Normalstimmen empfindlich sind, können die genauen Parameterwerte berechnet werden, die die Normalstimme bestimmen. Es wurde festgehalten, dass die Normalstimme in zwei Untergruppen zerfällt die jeweils für sich quantifiziert wurden.

Somit kann die Normalstimme als Norm-Referenz bei Stimmklassifikationen und Stimmbeurteilungen herangezogen werden.

In der Regel ist jedoch nicht der gesamte hier untersuchte Parametersatz für eine Stimmuntersuchungen erforderlich. Um die Anzahl der zu bestimmenden Größen zu begrenzen, wurde mittels der Fischerschen Diskriminanzanalyse eine Rangliste der Stimmparameter erstellt, in der die einzelnen Messgrößen nach ihrer Signifikanz für die Gruppentrennung sortiert sind.

## Literatur

- [1] Kiosses, Fröhlich, Michaelis, Strube, Kruse: Untersuchung von Stimmstörungen mit Kohonen-Karten Akt. phon.-pädaudiol. Aspekte; Band.: 7, Seite 73, (1999/2000).
- [2] Michaelis, Fröhlich, Strube: Selection and combination of acoustic features for the description of pathologic voice; J. Acoust. Soc. Am. 1998
- [3] Boianov, Chollet: Analysis of Pathological Voice; Res. Proj. ERB-CIPA-CT-92-0170 Paris