Phonetik I: Akustische Phonetik (V)

Inhalt: Akustische Phonetik

- Physikalische Grundlagen
- Akustogenese: Generierung des Sprachschalls im Sprechtrakt
- Regeln zur Modifikation von Formantfrequenzen
- Akustische Merkmale einzelner Lautgruppen
- *Akustische Analysemethoden* ←
- Einführung in das Arbeiten mit Sprachsignal-Analysesoftware
- Durchführung und Vorbearbeitung von Sprachschallaufnahmen
- Akustische Merkmale von Stimm- und Sprechstörungen

Hinweis:

-> aktuell und ausformuliert in: in MedAku\02\Analyse

Dieses Kapitel in dieser Veranstaltung nicht detailliert besprechen !!!

Akustische Analysemethoden

- Kurzzeitspektralanalyse (Fouriertransformation, Fensterung, DFT, FFT, Optimale Breit- und Schmalbandsonagramme)
- Methoden der spektralen Glättung (Cepstrum, LPC)
- Formantanalyse
- Grundfrequenzanalyse
- Intensitätsanalyse

Kurzzeitspektralanalyse

Erinnerung: Sprachsignale sind *zeitveränderlich* (nicht stationär);

Damit ist gemeint: Die Änderungen aufgrund der Artikulationsbewegungen.

- Die Frequenzanalysen sollen die Zeitveränderlichkeit von Sprachsignalen (die Artikulation und damit die Zeitverläufe der Formantfrequenzen) widerspiegeln:
- Jede einzelne Frequenzanalyse wird über ein kurzes Zeitintervall (ca. 10ms; 5-50ms) durchgeführt. In diesem Zeitintervall (Zeitfenstern) ist das Signal „quasistationär“

Wir unterscheiden: Frequenzanalyse generell (unendlich langes Zeitsignal) und

-> **Frequenzanalyse eines kurzen Zeitfensters:**
Kurzzeitspektralanalyse (Kurzzeitspektrum, Spektrum)

Spektralanalyse / Frequenzanalyse: Das Spektrum

Fouriertheorem: Jedes Signal kann eindeutig in Sinus- (und Cosinus-)Schwingungen mit definierter Amplitude und Phase zerlegt und umgekehrt aus den Amplituden- und Phasenwerten auch wieder eindeutig zusammengesetzt werden. (-> *Fouriertransformation*)

- Fourier*analyse*: Zerlegung eines Zeitsignals in Sinusschwingungen PM_044
- Fourier*synthese*: Zusammensetzung eines Signals aus Sinusschwingungen

Im Folgenden betrachten wir den Spezialfall:

- digitalisiertes Signal -> diskrete (= nicht kontinuierliche) Zeitpunkte und Zahlenwerte
- Kurzzeitspektralanalyse -> Fourieranalyse über ein endliches Zeitintervall

Fouriertransformation -> diskrete Fouriertransformation (DFT)

- Zeitintervall bestehend aus N Zahlenwerten -> N Zeitwerte x(n)
- Die DFT liefert N/2 Amplitudenwerte f(n) und N/2 Phasenwerte φ(n)
  (Erinnerung: Fouriertransformation bedeutet immer Frequenz- und Phasenanalyse)

Für uns interessant: das Amplitudenspektrum. Also:

- Analyse über das Zeitintervall $T = N * \Delta t$ ($fs = 1 / \Delta t$ Samplingfrequenz, Abtastrate)
  -> N Zeitwerte x(n) liefert:
- N/2 Amplitudenwerte / Frequenzwerte f(n) über das Frequenzintervall 0 ... fs/2 (Erinnerung: Abtasttheorem!)

Damit gilt für die Frequenzauflösung der DFT: $\Delta f = fs / N$

Also:

Die Frequenzauflösung wird bei konstanter Abtastrate um so feiner, je länger das Analysefenser ist.

Anmerkung: Die Samplingrate sollte für Formantanalyse ca. 10 kHz betragen, für Analyse auch der Rauschanteile mindestens 20 kHz.

Anmerkung: Wegen der Zeitveränderlichkeit des Sprachsignals sollte die Länge des Analysefensters auf ca. 10 ms begrenzt sein.

Anmerkung: CD-Samplingrate

Die Samplingrate 44.1 kHz wird gerne gewählt (→ CD)

Wegen der Formel $\Delta f = fs / N$ folgt:

Hohe Abtastrate → schlechte Frequenzauflösung

Für Sprachanalyse reicht eine Samplingrate von ca. 20 kHz (22.05 kHz) eigentlich.

(Bei Analyse von Stimmparametern: jitter etc. ist 44.1 kHz aber besser)

Sprachsignalanalyse:

Um bei 44.1 kHz eine gleiche Frequenzauflösung wie bei 22.05 kHz zu bekommen, muss die Zahl N bei der DFT / FFT verdoppelt werden.

→ Das Zeitintervall bleibt bei Verdopplung der Samplingrate dann gleich.

Der Rechenaufwand für die Durchführung der Analysen steigt damit aber erheblich. (ca. quadratisch)

Das ist aber bei den heute üblichen Rechnern nicht mehr spürbar.

Die Fast-Fourier-Transformation (FFT)

FFT ist ein Spezialfall der DFT (Diskrete Fouriertransformation)

Hintergrund: Effizienz von Algorithmen zur Fouriertransformation:

Wählt man die Fensterlänge N in Potenzen von 2, so kann die Rechenzeit der Fouriertransformation aufgrund geschickter Algorithmen *drastisch* reduziert werden:

Bestimmend für die Rechenzeit: Die Anzahl der Multiplikationen

Rechenzeit einer DFT:	$N * N$		$(= N^2)$				
Rechenzeit einer FFT:	$N * \log_{10} N$						
FFT's arbeiten mit N =	32,	64,	128,	256,	512,	1024,	2048
FFT spart Faktor:	10				100		1000

Achtung: Im Frequenzbereich wird nur die halbe Punktzahl einer DFT / FFT dargestellt: Der Frequenzbereich der DFT / FFT reicht bis zur Grenzfrequenz (fs/2)!

Je nach Samplingrate (ca. 10 bis 40 kHz) liegt das Zeitfenster im Bereich von ca. 5 bis 100 ms

Beispiel: N=128 und fs=10kHz (100 Punkte stellen 10 ms dar) -> T = 12,8ms

Die Fensterfunktion

Was bedeutet „Fensterfunktion“ und warum benötigt man diese überhaupt?

Beispiel: Analyse einer Sinusschwingung (eines reinen Tons)

- „theoretische“ Fouriertransformation liefert eine „Linie“: Unendlich genaue Frequenzanalyse aufgrund eines unendlich ausgedehnten Analysefensters

  Bsp.: hess_gsv3f_53ff: theoretische Wiederholung des Zeitsignals unendlich oft

  (Achtung: hier wird aber bereits ein endliches Analysefenster gewählt)

- die „theoretische“ Fouriertransformation entsteht in der Praxis aber nur, wenn die Länge des Analysefensters genau mit der Periodendauer des zu analysierenden Signals übereinstimmt.

  In der Realität ist das aber fast nie der Fall

  Beispiel: hess_gsv3f_59f (Amp.-, Freq.-Achse linear; Amp. log.; Freq. log.)

„reale“ Fourieranalyse:

- „Linienverbreiterung“ aufgrund der Endlichkeit des Analysefensters (endlicher Signalausschnitt zur Berechnung Analyse): $\Delta f = fs / N$
- Es entstehen *Artefakte* aufgrund des willkürlichen Beginns und Endes des Analysefensters:

  → Verbreiterung der Spektrallinie hess_gsv3f_59f

Hintergrund:

Das Analysefenster selbst kann als Rechtecksignal angesehen werden und hat ein eigenes Spektrum.

Jede Sinusschwingung im Fenster ist in der Praxis eine Überlagerung (Faltung) der Spektrallinie der Sinusschwingung mit dem Spektrum des Analysefensters.

Also: Aufgrund des „Spektrums des Analysefensters“ „verschmiert“ die Spektrallinie jeder Frequenzkomponente des zu analysierenden Signals. → Verbreiterung der Spektrallinie / Nebenmaxima

Weiterer Nachteil: Die Amplitude nimmt nur langsam ab (6db/Oktave)

hess_gsv3f_59f (-> eine Gerade in der Darstellung log Amp / log Freq)

Abhilfe:

Wir ersetzen das Rechteckfenster (= abrupter Anfang und abruptes Ende des Analysefensters) durch eine „Fensterfunktion“: hess_gsv3f_62

Durch die Fensterfunktion wird ein „sanfter Übergang“ vom restlichen Zeitsignal zum Analysebereich geschaffen.

Beispiele: (Hanningfenster, Hammingfenster, Kaiserfenster)

Aufgrund des sanften Übergangs sind die Amplitudenwerte der Nebenmaxima der Fensterfunktion verglichen zum Rechteckfenster geringer.

Bsp.: hess_gsv3f_63ff: immer ohne / mit Fensterfuntion

- Kosinusfunktion, ungedämpft; 5 Perioden hess_gsv3f_63f
- Kosinusfunktion, ungedämpft; 5,5 Perioden hess_gsv3f_65f besser mit Fensterfkn
- Kosinusfunktion, gedämpft; 5 Perioden hess_gsv3f_67f

Eigenschafen von Fensterfunktionen

- *Rechteckfenster*: erstes Nebenmaximum bei –16 dB (-> bei mehr als 16 dB Dynamikbereich verdecken sich die Spektralanteile gegenseitig)
- *Hanningfenster*: erstes Nebenmaximum noch recht hoch (-32 dB), aber die Nebenmaxima fallen danach sehr schnell ab. (siehe obige Beispiele)
- *Hammingfenster*: erstes Nebenmaxima recht tief (-42 dB), aber die weiteren Nebenmaxima fallen nur noch sehr langsam weiter ab.

Bei einer komplexen Spektralanalyse werden somit *die wahren Frequenzanteile des zu analysierenden Signals mit Fensterfunktion besser hervorgehoben* als ohne (= Rechteckfenster)

Allerdings führt die Fensterfunktion zu einer *Kürzung der effektiven Fensterlänge* -> Die Hauptmaxima sind gegenüber dem Rechteckfenster etwas breiter.

Also: Die Frequenzauflösung wird durch eine Fensterfunktion verringert.

Andererseits: Die reale Frequenzauflösung wird durch die hohen Nebenmaxima des Rechteckfensters in der Praxis noch stärker verringert.

Auswirkung von Fensterfunktionen bei realen Sprachsignalen: Vokale

Beispiele

- Vokal [ɛː], Analysefensterlänge 5 Perioden hess_gsv3f_72f
  - Teiltöne treten mit Fensterfunktion nicht klar hervor
  - Teiltöne treten nur in dem (praktisch unerreichbaren) Fall der Fensterlängen- und Perioden-Koinzidenz klar hervor

- Vokal [ɛː], Analysefensterlänge 4,5 Perioden hess_gsv3f_76f
  - Teiltöne treten nie klar hervor (ohne und mit Fensterfunktion; etwas besser mit)
  - Teiltöne liegen genau zwischen den Fourierwerten (Fourierwerte sind definiert durch die Fensterlänge!)

Lösung:

Heraufsetzen der Frequenzauflösung durch Nutzung einer längeren FFT über dem (kurzen) Analyseintervall → Einschieben von „Nullsignal“

Beispiel:

Vokal [ɛː], Analysefensterlänge 5 Perioden hess_gsv3f_74f

Vokal [ɛː], Analysefensterlänge 4,5 Perioden hess_gsv3f_78f

Ergebnis:

- Die Teiltöne treten nun klar hervor
- Die Absenkung der Nebenmaxima (zwischen den Teiltönen) durch die Fensterfunktion wird klar sichtbar

Optimale Breit- und Schmalbandsonagramme

Einführung:

Die Periodizitätsinformation kann aufgrund der Unschärferelation entweder im Zeitbereich oder im Frequenzbereich dargestellt werden: hess_gsv4f_21f

- Zeitbereich → Glottisimpulse; Breitbandsonagramm hess_gsv4f_23f
- Frequenzbereich → Teiltöne; Schmalbandsonagramm hess_gsv4f_25f

Zur Berechnung von Sonagrammen

Die Berechnung von Sonagrammen geschieht anhand einer zeitlichen Abfolge von FFT‘s

Frage nach der: hess_sv34_15

- optimalen *Schrittweite* (Parameterabtastintervall)
- der optimalen *Fensterweite* (d.h. die optimale FFT-Länge)
- der optimalen *Fensterfunktion* für eine Sonagramm

Darüber hinaus: Das Problem der *Amplitudenwichtung*:

Zur guten visuellen Darstellung eines Sonagramms (Grauwertdarstellung) ist das Abschneiden von Spitzen-Amplitudenwerten (clipping) hilfreich!!

Also: Oberhalb eines bestimmten Amplitudenwertes ist die Darstellung ganz geschwärzt;

unterhalb eines bestimmten Amplidudenwertes (vor dem Rauschsignal) ist die Darstellung ganz weiß

Dazu gibt es wenig Literaturdaten!

Erfahrungswerte für Breitbandsonagramme (→ siehe die Standardeinstellungen in Praat www.praat.org)

Methoden der spektralen Glättung

Spektrale Glättung ist die Basis für Formantanalyse.

Formantanalyse setzt spektrale Glättung voraus.

Hintergrund: Quelle-Filter-Theorie: hess_gsv4f_02ff PM_051

Das abgestrahlte Schallsignal ist immer eine Überlagerung von Quelle und Filter

Im Frequenzbereich: Überlagerung Teiltöne der Quelle und Filterfunktion

Im Grunde ist für die Formantanalyse nur die Filterfunktion wichtig. hess_gsv4f_05ff

→ Ich möchte die im Spektrum letztlich störenden „Teiltöne“ eliminieren.

→ „Glättung“ des Spektrums

Verfahren zur spektralen Glättung

Zwei wichtige Verfahren: Cepstrum und LPC

Zuvor: Erinnerung: **Fourieranalyse:**

Die Fourieranalyse überführt ein Signal in eindeutiger Weise vom Zeitbereich in den Frequenzbereich (Amplituden- und Phasenwerte).

„Eindeutig“ bedeutet, dass das Signal bei „Rücktransformation“ identisch entsteht.

Dies gilt allerdings nur bei Berücksichtigung von Amplituden- und Phasenspektrum.

Im Bereich der Sprachsignalverarbeitung wird zumeist nur das *Amplitudenspektrum* bzw. das *Leistungs(dichte-)spektrum* genutzt. Darstellungsmöglichkeiten:

- Amplitudenspektrum: Amplitudenbeträge |A|
- Leistungsdichtespektrum: A^2 oder logarithmisch $L = 10 \log A^2 = 20 \log |A|$ [dB]

Vorteil der logarithmischen Darstellung , auch genannt „Schall(druck)pegel“:

Da bei akustischen Schallsignalen sehr unterschiedliche Amplitudenwerte auftreten können (12 10er Potenzen), reicht die lineare Skala nicht zur Darstellung sehr kleiner Amplituden.

Der Logarithmus (die Schallpegelskala) führt zu einer *Kompression* dieser Skala und macht damit den ganzen Amplitudenbereich gut sichtbar.

(Die Schallpegelskala hat sich auch deshalb so gut durchgesetzt, weil das Gehör näherungsweise Schallamplituden ebenfalls logarithmisch verarbeitet.)

→ Amplitudenspektren werden in der Regel als Schall(druck-)pegel vs. Frequenz angegeben.

Cepstrum und Autokorrelationsfunktion (AKF)

Cepstrum: wichtiges Verfahren der spektralen Glättung: hess_gsv4f_12f

AKF: hess_sv4_14 AKF ist eine Möglichkeit zur F0-Bestimmung, steht aber in engem theoretischen Zusammenhang zum Cepstrum (und wird deshalb hier erwähnt)

Cepstrum und AKF sind spezielle Signaltransformationen in den Frequenzbereich und wieder zurück in den Zeitbereich: (FT = Fouriertransformation)

- AKF: FT -> Leistungsdichtespektrum A^2 -> inverse FT
- Cepstrum: FT -> logarithmisches Leistungsdichtespektrum -> inverse FT

Interpretation von AKF und Cepstrum

Wie kann man nun AKF und Cepstrum interpretieren?

hess_gsv4f_12f hess_sv4_14f

Zunächst: Was ist eigentlich eine Fouriertransformation?

Analyse eines Signals auf Sinus- (und Cosinus-)Anteile. Also:

Analyse des Signals auf seine „Welligkeit“:

- „lange“ Wellen -> niederfrequente Anteile
- „kurze“ Wellen -> hochfrequente Anteile („eckiges“ Signal)

In der gleichen Weise führt eine inverse FT eine Analyse der „Welligkeit“ des Spektrums durch.

Beispiel: Vokalspektrum: hess_gsv4f_05

- die kürzeste Welligkeit des Spektrums entsteht durch die Teiltonstruktur
- eine längere Welligkeit des Spektrums entsteht durch die Übertragungsfunktion, durch die Formanten

Cepstrum

Wichtiges Verfahren der spektralen Glättung: hess_gsv4f_12f

- Wir erkennen einen prominenten Reak im Cepstrum:

  Erstes Maximum im Cepstrum oder bei AKF ist ein Schätzwert für die Grundfrequenz (d = 1/f0) (wichtig für F0-Analyse siehe später)

- Abschneiden unterhalb dieses Peaks (-> Tiefpass) und Rücktransformation liefert das geglättete Spektrum

Also: Eine spektrale Glättung mittels Cepstrum passiert durch Tiefpassfilterung des Cepstrums (Liftering der Verzögerungswerte) und erneuter Fouriertransformation

Hess_sv5_23

Weitere Erläuterungen zum Cepstrum:

- Die x-Achse des Cepstrums: ist (nach Hin- und Rück-Fouriertransformation) eine Zeitachse. Die Zeitwerte heißen auch: „Verzögerungswerte“ d („delay“); es sind nicht nicht absolute Zeitwerte t („time“)
- Filterung im Cepstrum wird auch als „liftering“ bezeichnet: Hess_sv5_22

Aus dem geglätteten Spektrum können nun einfach die Frequenzen der Maxima der Übertragungsfunktion des Vokaltraktes abgelesen werden -> Formantfrequenzen
hess_gsv4f_05

Probleme bei der spektralen Glättung mittels Cepstrum:

Oftmals werden nicht alle Formanten erkannt:

Der Grund:

- [a, o, u]: F1 und F2 sind sehr eng benachbart → können nicht gut separiert werden
- [i]: F2 und F3 sind sehr eng benachbart → können nicht gut separiert werden
- F1 und F0 sind sehr nah (hohe Sprecherstimme): dann kann die Glättung nach dem Prinzip des „Ausfilterns bestimmter Verzögerungsbereiche“ nicht gut funktionieren.

Lösung: Das Prinzip der linearen Prädiktion.

Lineare Prädiktion (Linear predictive coding LPC)

Betrachtung im Frequenzbereich:

Lineare Prädiktion ist eine *Modellierung* der *mittleren Welligkeit* des Spektrums:

Unterscheidung:

- *Mittlere Welligkeit*: Welligkeit aufgrund der Formanten, aufgrund der Übertragungs-funktion des Ansatzrohres PM_051
- *Niedrige Welligkeit*: spektraler Abfall des Quellspektrums und durch Abstrahlung
- *Hohe Welligkeit*: aufgrund der Grundfrequenz und der Teiltöne

Das Modell hinter der LPC

Der LPC- Methode liegt ein einfaches Modell zur Übertragungsfunktion des Ansatzrohres zugrunde:

Annahme: Das Ansatzrohr kennt nur Resonanzen, keine Antiresonanzen

(Dies gilt, wenn die Schallanregung immer nur an der Glottis passiert, wenn die Schallabstrahlung nur über den Mund erfolgt und wenn der Nasenraum akustisch nicht angekoppelt ist. PM_050)

Es wird mit dieser Annahme ein spezielles Modell für die Übertragungsfunktion des Ansatzrohres angenommen:

ein *nur-Polstellen-Modell* (= ein „reinrekursives“ Filter, ein „Prädiktorfilter“). → Theorie der digitalen Filter siehe Vorlesungsskripte Hess

Aufgabe des LPC-Algorithmus ist nun die Berechnung einer optimalen Übertragungsfunktion des Ansatzrohres. Hess_sv5_02ff

Dies geschieht durch Anpassung der Filterfunktion an das Signalspektrum, so dass der auftretende Fehler (Differenz zwischen Filterfunktion und Signalspektrum) möglichst gering ist: hess_sv34_21

Das Ergebnis ist eine Menge von LPC-Koeffizienten.

Die LPC liefert:

- Übertragungsfunktion (Impulsantwort) definiert durch die Lage der Pole (LPC-Koeffizienten)
- „Fehlersignal“ / „Residualsignal“: dieses Signal entspricht dem Quellsignal und ergibt nach Faltung mit der Impulsantwort das Ausgangssignal der LPC-Analyse (das vom Mund abgestrahlte Schallsignal)

Anzahl der LPC-Koeffizienten

Achtung: Die Anzahl der LPC-Koeffizienten ist frei wählbar!

Die Anzahl N der LPC-Koeffizienten ist definiert durch die Anzahl der Pole (der Formanten), die dargestellt werden sollen. hess_sv34_21

Faustregel: ein Formant pro 1000 Hz; 2 Koeffizienten pro Formant → N = 10 bei fs = 10 kHz (f_{grenz} = 5 kHz)

Bsp.: Hess_sv5_06:

Der Fehler sinkt drastisch bis ca. 10 Koeffizienten (bei 10kHz Abtastrate), weil ab hier die spektrale Einhüllende gut modelliert ist.

Hess_sv5_07ff

von links: Fehlersignal, Spektrum des Fehlersignals, Filterfunktion, Impulsantwort, Verteilung der Pole

Faktor μ: Differenzierung des Signals → Anhebung der hohen Frequenzen (μ → 1)

Methoden zur Berechnung der LPC-Koeffizienten

Es gibt mehrere unterschiedliche Methoden und es gibt unterschiedliche Darstellungen der Koeffizienten (Methoden: Autokorrelations- und Kovarianzmethode)

Autokorrelationsmethode („Stationärer Ansatz“: „Begrenze das Signal“ → Fensterung)

Das zur Analyse nötige Signalfenster ist bei dieser Methode lang;

Nichtstationäres Signal: Lösung wie bei FFT: Das Signalfenster wird durch 0-Werte ergänzt.

Vorteile dieser Methode:

- garantierte Stabilität des Filters → Robustheit
- Verfügbarkeit der schnellen PARCOR-Methode (Rekursive Berechnung der Prädiktorkoeffizienten)

Nachteil: Hess_sv5_10f

- Oft erhebliche Fehler in der Formantbandbreite (aufgrund des langen Fensters)

Kovarianzmethode („Nichtstationärer Ansatz“: „Begrenze die Meßmethode“)

Es werden nur kleine Signalausschnitte genutzt; es ist keine Fensterung nötig.

Vorteile: Hess_sv5_10f

- Liefert wesentlich genauere Ergebnisse als die Autokorrelationsanalyse
- Benötigt nur kurzen Signalausschnitt → ist geeignet zur periodensynchronen („pitch-synchronen“) Analyse

Nachteil:

- Das berechnete Filter kann instabil sein (Pole außerhalb des Einheitkreises)

Anmerkung:

Die PARCOR-Methode liefert in einer Formulierung direkt die Reflexionskoeffizienten des Ansatzrohres und kann damit aus dem akustischen Signal Querschnittsflächen des Ansatzrohres berechnen.

Die Methode ist mäßig erfolgreich: Hess_sv5_18

Ein Problem ist theoretisch nicht mögliche Separation von Quelle (inkl. Klang) und Filterfunktion hess_gsv4f_02f

Formantanalyse

Messung der Formantfrequenzen (F1, F2, F3, ...) über den Zeitverlauf einer Äußerung; manchmal auch der Formantbandbreiten oder –amplituden

Definition: Formanten = Resonanzen des Vokaltraktes (Ansatzrohres), also:

- Maxima des (geglätteten) Kurzzeitamplitudenspektrums , oder:
- Pole der Übertragungsfunktion des zugrundeliegenden Vokaltraktmodells

Fragen:

- Wie funktioniert eine Formantanalyse?
- Welche unterschiedlichen Verfahren gibt es?
- Was sind die optimalen Parameter-Einstellungen für eine Formantanalyse? (Fensterlänge, Fensterfunktion, ...)

Das Grundproblem: (gilt auch für F0-Analyse)

Während ein Sonagramm, eine FFT, eine Intensitätsanalyse, eine Signalfilterung (etc.) ohne das Fällen von Entscheidungen berechnet / ausgeführt werden kann (nicht interpretative Analyse), müssen die Formantwerte (und auch die Grundfrequenz) anhand des Signals „ausgewählt“ werden (interpretative Analyse; Entscheidungen sind zu treffen).

Manchmal gibt es mehrere „Kandidaten“ für F0 (Obertöne);

Manchmal liegen z.B. F1 und F2 so nah zusammen, dass sie nicht „aufgelöst“ werden können; ....

→ *Formantanalyse ist eine „interpretative Analyse“:* Es müssen Parameter der Produktion anhand des Sprachsignals „zurückgerechnet“ werden.

Formantanalyse gliedert sich prinzipiell in 2 Teile:

- *Bestimmung von Kandidaten*: alle Frequenzen, die für F1, F2, F3, ... in Frage kommen nur anhand des momentanen Analysefensters

  Verfahren:

  - *Peak-Picking*: Gewinnung der Maxima anhand der geglätteten Spektren (evtl. Interpolation auf der Frequenzachse zur Verbesserung der Frequenzauflösung (z.B. 3-Punkt-Interpolation)

    Zur Glättung wird hier oft auch LPC-Methode eingesetzt, dann Peak-picking über das berechnete LPC-Spektrum ausgeführt (z.B. McCandless 1974)
  - *Root-solving*: Berechnung der Frequenzen der Pole (der Maxima der ÜF) direkt aus den Koeffizienten der LPC
- *Selektion von Fi-Werten* anhand von „vertikaler“ und „horizontaler Information“
  - Vertikale Information: ein Formant liegt ungefähr alle 1000 Hz; allerdings je nach Sprecher (Mann, Frau, Kind) modifiziert
  - Horizontale Information: Der Zeitverlauf der Formanten ist stetig (d.h. die Formanttransitionen sind stetig).

Generell ist Formantanalyse aber ein interpretatives Verfahren und damit unsicher:

Es kann nie mit 100%iger Sicherheit davon ausgegangen werden, dass der vorgeschlagene Frequenzwert auch tatsächlich den entsprechenden Formaten repräsentiert.

Problempunkte:

- unzureichende spektrale Auflösung im Falle zweier nah benachbarter Formanten (z.B. Eckvokale: F1 und F2 bei [a] und [u]; F2 und F3 bei [i])
- Formanten sind bei Nasalen durch benachbarte Antiformanten „eingeebnet“ und damit nicht gut erkennbar
- Formanten können durch Nullstellen im Hüllspektrum des Anregungssignals kompensiert werden
- Höhere Formanten sind generell sehr schwach. Es sollte deshalb nicht oberhalb von F4/F5 gemessen werden. Manchmal ist schon F3 sehr schwach (z.B. bei tiefem F1 und F2; z.B. [o] und [u])

Wegen dieser Fehleranfälligkeit (der Problematik hinsichtlich der Zuverlässigkeit) von Formantanalyse-Algorithmen werden Formantsynthesen zumeist mit „robusten“ Parametern der Vokaltraktübertragungsfunktion (z.B. mit den LPC-Koffizienten selbst) durchgeführt.

- Es wird auf die explizite Extraktion von Formantwerten verzichtet.
- Die LPC-Parameter parametrisieren die geglättete Übertragungsfunktion aber vollständig und sind somit auch eine hinreichende Darstellung der ÜF.

Im Bereich der klinischen Phonetik ist es daher unvermeidlich, mittels maschineller Verfahren gewonnene Formantdaten visuell zu kontrollieren und ggf. manuell zu korrigieren.

Beispiele für Formanterkennungs-Algorithmen:

- Algorithmus von McCandless 1974 (auf LPC-Basis) Hess_sv5_28
- Verfahren von Atal und Schroeder 1978 (auch auf LPC-Basis; aber Berechnung von Formanten und Antiformanten) Hess_sv5_36

→ Formantextraktion ist in der Regel hochkomplex

Grundfrequenzanalyse

Fragen / Themengebiete:

- Wie funktioniert eine F0-Analyse?
- Welche unterschiedlichen Verfahren gibt es?
- Wie funktioniert die Messung weiterer Parameter der Anregung (Jitter, Shimmer und HNR)?

Ablauf einer Grundfrequenzanalyse

- Zunächst: Stimmhaft / Stimmlos – Detektion (z.B. anhand von Nulldurchgangsdichte); Frage: Liegt überhaupt ein stimmhaftes Signal vor? (Vorverarbeitung) hess_sv4_04
- Durchführung der F0-Erkennung im Bereich stimmhafter Signalabschnitte hess_sv4_07 (hier Vokal [ɛː]; unterschiedliche Register)

Es existieren viele (hunterte) Methoden / Algorithmen; aber keine funktioniert „einwandfrei“ → visuell-auditive Überprüfung und manuelle Korrekturen sind oft unumgänglich

Grundproblem der F0-Analyse (wie bei der Formantanalyse):

F0 (die Grundfreqeunz) ist letztlich ein *Modell*-Parameter der Sprachproduktion (Schwingungsfrequenz der Simmlippen) und kann daher nur indirekt anhand des akustischen Signals erschlossen werden.

Praktische Probleme bei der Grundfrequenzanalyse:

- Die Grundfrequenz kann von Periode zu Periode schwanken (Larynx = biologisches System) (gerade im Fall pathologischer Stimmen) → Unregelmäßigkeiten im F0-Verlauf
- Das Sprachsignal ist vom Vokaltrakt überformt und deshalb kann das abgestrahlte Signal von Periode zu Periode der Grundfrequenz anders aussehen
- Die Grundfrequenz (bei nicht bekanntem Sprecher) kann über eine weite Frequenzspanne variieren: 50-800 Hz
- Die Grundfrequenz kann in den Bereich des ersten Formanten geraten (insbesondere bei hohen Vokalen [i], [u]). Das führt zu Problemen bei der automatischen Analyse
- Das Signal kann durch schlechte Übertragungsstrecken verzerrt sein (z.B. Telefon, schlechte analoge Aufnahme, ...)

F0-Algorithmen

Prinzipiell 3-stufig:

- Vorverarbeitungstufe: Signal/Null- und stimmhaft/stimmlos-Erkennung hess_sv4_04
- Extraktionsstufe: Liste von F0-Kandidaten erstellen; Folge von F0-Schätzwerten (und Vorauswahl eines wahrscheinlichsten Kandidaten)
- Nachbearbeitungsstufe: Fehlerkorrektur durch horizontalen Vergleich Werten mit benachbarten Analysefenstern; evtl. Glättung des gesamten F0-Verlaufes

Arten von Algorithmen: hess_sv4_15

- Frequenzbereichsalgorithmen / Algorithmen mit Durchführung einer Kurzzeitanalyse (nicht unbedingt Kurzzeitspektralanalyse)
- Zeitbereichsalgorithmen (ohne Kurzzeitanalyse)

Algorithmen unter Nutzung der Kurzzeitanalyse

Generell bei allen Verfahren: Länge des Analyseintervalls länger als bei Vokaltrakt-Analyse: ca. 2-3 Grundperioden (länger oder kürzer ist suboptimal) hess_sv34_15

Methoden: hess_sv4_15

- Autokorrelationsfunktion: AKF: liefert Maximum bei T = 1 / F0 hess_sv4_14

  AKF ist die zeitverzögerte Multiplikation der Signalwerte des Fensters

  AKF ist die Rücktransformation des Leistungsdichtespektrums (siehe oben)

- Distanzfunktion AMDF (average magnitude difference funktion): liefert Minimum bei T = 1 / F0 hess_sv4_15

  AMDF ist die Summe der zeitverzögerten *Betragsdifferenzen* der Signalwerte des Fensters (AKF der zeitverzögerten Amplitudenquadrate)

  AMDF hat keine direkte Entsprechung im Frequenzbereich

- Cepstrum: liefert Maximum bei T = 1/F0 hess_sv4_15 hess_sv4_17 hess_sv4_18

  Cepstrum ist die Rücktransformation des logarithmierten Leistungsdichtespektrums (Vorsicht: hier treten auch tiefer bereits hohe Peaks auf → siehe Formantanalyse)

Ein komplexes Beispiele für F0-Bestimmung mittels Kurzzeitspektralanalyse:

Der Algorithmus von Martin (1981) hess_sv4_16

- Fouriertransformation bei Grenzfrequenz 2kHz (schnell)
- Weitertransformation mittels Kammfilter → Idee: Teiltöne sind auf der Frequenzskala gleich separiert

Vorteile dieser Methoden:

Unempfindlichkeit gegen Phasenverzerrungen, Rauschen, Bandbegrenzungen bei tiefen Frequenzen

Zeitbereichsalgorithmen (ohne Kurzzeitanalyse)

Es werden Periodengrenzen (Markierer) gesetzt: Es werden also im Zeitbereiche jeweils der Periodenanfang bzw. das Periodenende definiert.

Der zeitliche Abstand T zwischen 2 Markern definiert F0 = 1/T

Es wird die Tatsache genutzt:

- Die Luftschwinugungen im Vokaltrakt werden zum Verschlusszeitpunkt der Stimmritze maximal angeregt
- Die Vokaltraktschwingungen (Formanten) klingen im Zeitbereich noch innerhalb einer Grundperiode merklich ab. (Amplitudenverringerung) hess_sv4_22

Einige einfache Algorithmen:

- Nulldurchgangsextraktionsstufe: hess_sv4_23 (unten)
  - Beseitigung des Einflusses der Formanten durch ein Tiefpassfilter
  - Extraktion jedes Nulldurchganges mit aufsteigender Flanke

  Problem: Der Tiefpassfilter kann nicht immer alle Formanten wegfiltern: benötige mindestens Filter mit 18 db/Okatve (Fehler → violett dargestellt)

- Schwellwertextraktionsstufe: hess_sv4_23 (Mitte)
  - Beseitigung des Einflusses der höheren Formanten durch Tiefpassfilter
  - Extraktion der Zeitpunkte, die einen definierten Schwellwert mit definiert gerichteter Flanke übersteigen

  Problem auch hier: der Einfluss der Formanten
- Schwellwertextraktionsstufe mit Hysterese: hess_sv4_23 (oben)
  - Beseitigung des Einflusses der höheren Formanten durch Tiefpassfilter
  - Extraktion der Zeitpunkte, die einen definierten hohen Schwellwert mit definiert gerichteter Flanke überschreiten, nachdem zuvor ein (niedrigerer) Schwellwert mit umgekehrt gerichteter Flanke durchlaufen wurde

  Dieser Algorithmus erlaubt somit eine Teilschwingung im Signal, reagiert aber dennoch empfindlich gegen das Vorhandensein höherer Formaten

Reale Zeitbereichs-F0-Algorithmen sind komplex:

Übersicht über die Zeitbereichsverfahren: hess_sv4_21

Beispiel: Unterdrückung von Peaks aufgrund von Hemmzeit und Abkling-Amplitude hess_sv4_26

Nachbearbeitung bei F0-Analyse

- Listenkorrektur: Eliminierung von Grobfehlern (z.B. Oktavfehlern); hess_sv4_29
  Dabei kann an „sicheren“ Signalstellen (z.B. im Silbenmittelpunkt) begonnen werden;
  Korrektur nach beiden Seiten hin zum (stimmlosen) Silbenrand
- Glättung: z.B. Tiefpassfilterung der erhaltenen Folge von F0-Werten
  Sollte erst nach Listenkorrektru vorgenommen werden, da ein „Ausreißer“ die Gesamtkurve in Richtung des Ausreißers bewegt. hess_sv4_30

Anmerkung zu Zeitbereichs-F0-Detektion:

Es können aufgrund der Schwellwertdefinition Artefakte auftreten: hess_sv4_24

Rekonstruktion des glottalen Luftstroms / der Stimmlippenschwingung

Geht über eine F0-Analyse hinaus

Erinnerung an die Sprachproduktion: hess_sv4_32

- Quellsignal: Teiltöne und deren Einhüllende → Stimmklang
  Dann: Überformung des Quellsignals durch das Ansatzrohr
- Achtung: „normale“ LPC-Analyse liefert nur die Periodizität
- *Inverse Filterung* liefert eine Rekonstruktion des Quellsignals hess_sv4_33

Das Signal nach inverser Filterung ist sehr geeignet zur F0-Analyse; Allerdings kann dieses Signal nur sehr aufwendig gewonnen werden.

Auf ähnlichen Ideen basierender Algorithmus von Hess (1994): hess_sv4_34

- Schätzung der Impulsantwort (IA) über LPC-Analyse PM_050
- Korrelation der IA mit dem Signal → Maxima am Zeitpunkt des Glottisverschlusses

Intensitätsanalyse

Relative Intensitätsanalyse: einfach realisierbar; Für phonetische Analysen zumeist ausreichend.

Absolute Intensitätsanalyse (= Schallpegelmessung) ist bereits relativ aufwendig;

Ist in der Phoniatrie aber z.B. zur Messung des Stimmfeldes eines Sprechers unvermeidbar.

Intensitätsanalyse anhand des digitalisierten Signals

Kurzzeiteffektivwert / Kurzzeitleistung (Kurzzeitenergie): Sampelwerte quadrieren, über das Fenster aufaddieren; durch Fensterlänge teilen; dann: Wurzelziehen hess_sv34_15

Kurzzeitamplitude: Beträge der Sampelwerte bilden, über das Fenster aufaddieren; durch Fensterlänge teilen.

(dieser Wert ist schneller / unaufwendiger berechenbar)

Problem: Je nach gewählter Fensterlänge wird der Intensitätsverlauf des Signals verschieden stark geglättet!