

WELCHE BEWEISE GIBT ES FÜR EINE NICHT LINEARE BEZIEHUNG ZWISCHEN DER PRODUKTION UND DER PERZEPTION DER SPRACHE?

Im Vortrag letzte Woche lernten wir, dass es bei der Artikulation Bereiche gibt, die von starker Variabilität gekennzeichnet sind, deren akustische Ergebnis allerdings wenig variabel ist. Umgekehrt gibt es Bereiche, wo eine nur kleine artikulatorische Änderung eine starke Veränderung in der Akustik bewirkt. Die Abbildung hier bildet diesen Zusammenhang nochmal ab. Kenneth Stevens beschreibt nicht allein diesen Zusammenhang, sondern auch jenen, der die Akustik mit der Perzeption in Relation stellt. Auch diese wäre mit derselben Abbildung darstellbar, man müsste nur die Achsenbeschriftung verändern. Auch kann man sich diese Abbildung als natürlich sehr schematisierte Darstellung zweier adjazenter, also direkt benachbarter Phone im Zeitsignal vorstellen, wenn X die Zeit abbildet, und Y wiederum die Akustik-Parameter.

(nächste Folie)

Die kontinuierlich ablaufenden Bewegungen der Artikulatoren erzeugen einen akustischen Output, der gegliedert ist. Diese Gliederung zeigt sich anhand abrupter Diskontinuitäten ("dislocations" im spektralen Muster). Diese sind mehr oder weniger stark ausgeprägte "Grenzsteine" (landmarks) an den Phon-Grenzen, welche besonders stark ausgeprägt sind an den Grenzen zwischen Vokalen (offene Passage im Ansatzrohr) und Konsonanten (Verengung im Ansatzrohr, am deutlichsten also bei Plosiven).

Die Mehrheit der Konsonanten wird durch die Verengung in der oralen Region des Vokaltraktes artikuliert. Diese Verengung ist ausreichend um bestimmte Diskontinuitäten im Spektrum des Signals zu verursachen. Beispiele für konsonantische Diskontinuitäten kann man anhand des Sonagramms des Satzes "The bike was Danish" sehen.

(nächste Folie)

Wenn man sich FIGURE1 vor Augen hält, so entspricht die Region II den Landmarks. An diesen Grenzen, den Phon-übergängen, gibt es offenkundig gewisse akustische Merkmale, die die in fließender Sprache auftretenden Laute von isoliert gesprochenen Lauten unterscheiden (-->Koartikulation und Steuerung). Diese akustischen Merkmale jedoch ermöglichen oft erst, die betreffenden Laute zu erkennen; (daher zieht Stevens eine Brücke zwischen diesen Signalparametern und den distinktiven Merkmalen der Phonologie.)

Diese Merkmale beanspruchen zwar einen beachtlichen Teil des an die landmark angrenzenden Signals eines bestimmten Phons, dennoch wird dieses

Phon (und ebenso das adjazente) eindeutig als Phon XYZ erkannt, d.h. der Hörer kann eindeutig zwei Laute von einander unterscheiden und abgrenzen (Trotz Koartikulation). Auch dies wird durch FIGURE1 ausgedrückt. Ein "Artikulator" (dies sind entweder die Lippen, das Zungenblatt oder der Zungenkörper) ist für die akustische Diskontinuität im Signal verantwortlich (principal articulator), beeinflusst das Signal also am meisten, indem er die "Grenzsteine" setzt. Dies gliedert das Signal in deutlich unterscheidbare Abschnitte. Die Abschnitte selbst jedoch sind auch nicht "steady state", sondern durch die adjazenten Abschnitte beeinflusst. Bleiben wir jedoch bei den Grenzen.

(nächste Folie)

Stevens unterscheidet drei Möglichkeiten, eine akustische Diskontinuität zu erzeugen:

1.: Unterschiedliche Arten der Primärschallerzeugung: z.B. Friktionsgeräusch/Plosivlösung (stärkere höhere Frequenzanteile) usw. vs. glottale Anregung (stärkere tieffrequente Frequenzanteile)= freie vs. verengte/verschlossene Passage im Ansatzrohr

2.: Orale vs. nasale Resonanz (Antiformanten)

3.: Nur laterale Passage im (oralen) Ansatzrohr vs. vollständig freie Passage (Liquid (/l/, engl./r/) vs. Vokal)

Diesen drei Möglichkeiten gemeinsam ist der Unterschied zwischen offener oraler Passage und einer Verengung durch einen der drei principal articulators. Bei jedem dieser Grenzsteine ist also mit einem Übergang zu/von einem Konsonanten zu rechnen. Der Artikulationsmodus und die Artikulationsstelle des Konsonanten lassen sich an den "acoustic cues" in der Umgebung der "landmark" ablesen und sind oftmals redundant kodiert:

(nächste Folie)

Beispiel1: Wie Phonetikern altbekannt ist, lassen die Formanttransitionen in den Vokalen in Richtung gewisser "Loci" auf die Artikulationsstelle des benachbarten Konsonanten schliessen. Auch die geräuschhaften Anteile des Konsonantensignals (z.B: Frikativrauschen, Plosionsgeräusch) geben mit ihrer Lage im Spektrum Hinweise auf die Artikulationsstelle (/d/ hat stärkere Amplitude im höherfrequenten Teil des bursts als /b/ wegen des noch nachgeschalteten Resonanzraumes)

Beispiel2: Nasale sind grundsätzlich durch wesentlich schwächer ausgeprägte Formanten gekennzeichnet. Ihre Lage ergibt sich aus der Artikulationsstelle (Verschlussstelle), die den Raum zwischen Glottis und

Verschlussstelle zum Resonator macht, dessen akustischer Output wiederum durch der Resonator Nase gefiltert wird, bevor der Schall die Nasenlöcher verlässt. An den Formanttransitionen in den adjazenten Vokalen ist ebenfalls die Artikulationsstelle kodiert, die Nasalität weitet sich oft auf die Vokale aus, so dass diese zumindest teilweise nasaliert sein können (Koartikulation als Redundanz?!) Im Gegensatz hierzu sind Vokale, die Laterale adjazent sind, sofort rein vokalisches. Lateral-Vokal-Übergänge sind gekennzeichnet durch ein starkes Ansteigen der Amplitude des zweiten Formanten

(nächste Folie)

Beispiel3: stimmhafte vs. stimmlose Frikative, die adjazent zu Vokalen sind: (divine vs. define)

*Unter den Sonogrammen sieht man F0 Kurve versus Zeit und Spektrum abgetastet an unterschiedlichen Stellen der Äusserung. Diese Stellen sind: vor der Konstriktion des jeweiligen Frikatives, circa 30 Millisekunden nach der Konstriktion des jeweiligen Frikatives und nach der Verschlusslösung. Jede von diesen Abbildungen enthält Information über die Stimmhaftigkeit des jeweiligen Frikatives.

Erstens sind dies Merkmale im Frikativsignal. Wie oben beschrieben, besitzen Frikative allgemein einen hochfrequenten Rauschanteil. Dieser ist bei stimmhaften Frikativen schwächer ausgeprägt; dafür besitzen diese einen tieffrequenten Anteil glottalen Ursprungs (der aber nicht über das gesamte Signal hinweg vorhanden sein muss). Dieser tieffrequente Signalanteil ist aber gegenüber demjenigen des Vokals um 5-15dB schwächer.

Zweitens beeinflusst aber auch die Tatsache, ob der Frikativ stimmhaft oder stimmlos ist, auch den adjazenten Vokal. Ist der Frikativ stimmlos, beginnt der Vokal mit höherer F0 (21 Hz in diesem Beispiel), als wenn der Frikativ stimmhaft ist; dies ist darauf zurückzuführen, dass die Stimmlippen nach einem stimmlosen Frikativ gespannter sind als nach einem stimmhaften. Zudem lässt das Spektrum des Vokals darauf schließen, ob der Frikativ davor stimmhaft oder -los war (wegen glottal spreading).

Im vorliegenden Beispiel sind die Formanten des dem stimmlosen Frikativ folgenden Vokals ausgeprägter in dem Sinne, dass sie eine höhere Amplitude, aber eben auch einen steileren Verlauf haben als der Vokal nach dem stimmhaften Frikativ. Unabhängig von der Stimmlage und anderen individuellen Charakteristika einzelner Sprecher ergeben sich wieder und wieder robuste Muster, die erstens das Sprachsignal in Segmente gliedern (mittels der "landmarks") und zweitens diese Segmente bestimmten Phonemen zuzuordnen erlauben (mittels der acoustic cues). Die Robustheit dieser Muster

ist ein weiterer Hinweis auf den quantalen Charakter gesprochener Sprache. Halle und Stevens sehen daher diese robusten Muster (die acoustic cues) als akustische Korrelate der distinktiven Merkmale innerhalb der Phonologie.

(nächste Folie)

Betrachten wir nochmals Stevens' Abbildung.

Wenn man die Verarbeitung von sprachähnlichen Signalen durch Menschen berücksichtigt, ergibt sich ein ähnliches Bild. In diesem Fall wird die Y-Achse der auditorischen Antwort entsprechen und die X-Achse entspricht dem akustischen Parameter. Das Ausmass der auditorischen Antwort zeigt eine non-monotone Änderung, wenn der akustischer Parameter manipuliert wird. Region II kann man als einen Grenzbereich betrachten im Sinne, dass wenn der akustischer Parameter sich in dieser Region ändert, die auditive Antwort sich von einem Muster zum anderen verändert.

Man vergleiche hierzu Professor Harringtons Abbildungen 1-3 der letzten Stunde. Sowohl für die Artikulatorisch-akustische Relation wie auch für die akustisch-perzeptive Relation gilt also (laut Stevens), dass es Quantalgebiete (Q) als auch Transitionsgebiete (T) (Stevens nennt diese auch "discontinuities" oder "landmarks") gibt. Da die Akustik das verbindende Glied zwischen Artikulation und Perzeption ist, ergibt sich die Kette Artikulation-->Akustik-->Perzeption; da an beiden Transformationen quantale Effekte behauptet werden, ist mit einer Verstärkung beider Effekte zu rechnen.

Prof. Harrington hat letzte Woche bereits das Drei-Rohr-Modell (Gekoppelte Resonatoren) vorgestellt und dargelegt, wie man damit Eigenfrequenzen (Formantlagen) berechnen kann (Verengungsstelle, Rohrlänge, Verengungsbreite). (Formanten oberhalb des F4 brauchen nicht in Betracht gezogen werden, da 1. die akustischen Verluste oberhalb des Frequenzbereichen von 3-4 kHz gross sind und 2. die Frequenzauflösung eines Menschen in diesem Bereich immer geringer wird, so dass zwar vielleicht messbare, aber nicht perzeptierbare Unterschiede im Signal festzustellen wären)

(nächste 2 Folien)

Sein letztes Beispiel verdeutlichte bereits den quantalen Charakter der Vokale [i] (F2 und F3 liegen nahe beieinander) und [u] (F1 und F2 liegen nahe beieinander). Hierzu Figure 7 und Figure 10 (Vier-Rohr-Modell wegen Lippenrundung: F1 ist tiefer als ohne Lippenrundung) Lippenrundung ergibt in jedem Fall einen akustischen Effekt (wg. des Vier-Röhren-Modells), der quantale Eigenschaften aufweist.

(nächste Folie)

Sie erlaubt eine stärkere Annäherung zweier Formanten, als dies ohne Rundung ohnehin zu erwarten wäre. Dies gilt insbesondere für F1 und F2 im Fall der hinteren Vokale, aber eben auch bei vorderen gerundeten Vokalen liegen die Formanten F2 und F3 noch näher beieinander als bei ungerundeten - und dies bei relativer Unsensibilität in Bezug zur Konstriktions-Position.

Der wahrscheinlich noch wichtigere Effekt der Lippenrundung ist jedoch vermutlich derjenige, dass der jeweils höhere eines Paares breitbandiger, aber von geringerer Maximalamplitude ist, so dass das Formant-Paar als "single-peaked-prominence" erscheint, im Gegensatz zur "two-peaked-prominence" ungerundeter Vokale. Dieses Phänomen wird noch für den Bereich der Perzeption eine bedeutende Rolle einnehmen. Doch dazu später...

(nächste Folie)

Turbulence noise als Quelle: Bekanntermassen können Friktionsgeräusche/Plosivlösungsgeräusche ebenso als rauschhafte Anregung des Resonators dienen. Eine Tatsache, die diese Art der Rohschallerzeugung quantal macht, ist die, dass an jenen Stellen, an denen unmittelbar vor der Rohschallerzeugungsstelle ein Hindernis (z.B. die Schneidezähne) sich befindet, dort der Rohschall eine um 20 dB höhere Amplitude aufweisen kann als derjenige an Stellen ohne zusätzliches Hindernis.

Ein weiterer quantal zu nennender Effekt ist folgender: wenn sich der Artikulator an die Artikulationsstelle annähert, setzt an einem bestimmten Punkt das Friktionsgeräusch ein. Bei weiterer Annäherung ist bald das Amplitudenmaximum erreicht. Auch wenn die Konstriktion nun noch weiter verengt wird, ändert sich die Amplitude des Rohschalles nicht mehr wesentlich. (Dies gilt ebenso für die Öffnung des Vokaltraktes). Dies bedeutet: es ist keine besondere Präzision bei der Konstriktionsbildung zur Erzeugung eines Friktionsgeräusches erforderlich. Dies gilt so nur für stimmlose Frikative, da bei stimmhaften die aerodynamisch-akustischen Vorgänge komplizierter sind und daher hierbei grössere Präzision erforderlich wäre, um über die gesamte Dauer des stimmhaften Frikativs Friktionsgeräusch und glottale Anregung gleichzeitig aufrechtzuerhalten. Zur Erkennbarkeit als stimmhafter Frikativ scheint es allerdings ausreichend zu sein, wenn der Frikativ nur teilweise mit glottaler Anregung produziert wird. Ist er erstmal entstimmt, wird das Friktionsgeräusch wieder leichter produzierbar (wegen der Quantalität stimmloser Friktionen) (dies könnte also die Erklärung sein, warum stimmhafte Frikative so häufig teilentstimmt produziert werden).

Ein weiteres Beispiel für einen abrupten Wechsel von einem Quantalgebiet ins nächste ist der Übergang [s]-[S]. Das völlig andere Spektrum für [S] erklärt sich aus dem Umstand, dass beim [S] im Gegensatz zum [s] ein Raum unter dem

Zungenblatt abrupt akustisch zugeschaltet wird, so dass der Resonanzraum vor der Primärschallerzeugung wesentlich grösser wird.

(nächste Folie)

Das eigentlich Interessante an der Frage nach der quantalen Natur der gesprochenen Sprache ist jedoch, was der Hörer wahrnimmt. Der Hörer transformiert schliesslich ein akustisches Signal in ein Perzept. Daher stellt Stevens psychophysikalische Experimente vor. (Identifikationstests, Diskriminationstests, similarity judgements).

Spektrale Prominenzen: Die Bandbreiten von Formanten nicht-nasaliertes, mit modal voice artikulierter Vokale sind für gewöhnlich schmäler als die "kritischen Bandbreiten" Zwickers (BARK). Im Gegensatz hierzu sind die Bandbreiten von nasalierten Vokalen und mit breathy-voice produzierter Vokale breiter als die Bandbreiten der auditorischen Filter. Dieses Phänomen wird uns noch begegnen.

(nächste Folie)

Nähe zweier Formanten: Pseudo-Vokale mit zwei Formanten werden synthetisiert; variiert werden die Entfernungen zwischen den Formanten (sowie später ihre relativen Amplituden). Ein einformantiger Klang ist der Vergleichsstimulus. Dessen Formant-Frequenz soll an die zweiformantigen Beispiele angeglichen werden:

(nächste Folie)

Relative Formant-Lage: Oberhalb eines kritischen Wertes (3.5 Bark) der Entfernung der beiden Formanten wird der Vergleichsformant entweder an den einen oder den anderen der beiden Formanten angeglichen (oder zufällig verteilt). Unterhalb des kritischen Wertes wird der Vergleichsformant so eingestellt, dass er zwischen den beiden Formanten liegt.

(nächste Folie)

Wird nun im Bereich unterhalb des kritischen Wertes der relativen Entfernung einer der beiden Formanten in der Amplitude gedämpft, so wird unterhalb eines kritischen Wertes der Amplitudendifferenz der Vergleichsformant zwischen die Formanten plaziert (wie bei denen mit identischer Amplitude), oberhalb des Wertes jedoch nur noch dem stärkeren angeglichen. Dies heisst, es gibt eine Schwelle bei der Amplitudendifferenz nah angrenzender Formanten, die bestimmt, ob noch zwei Formanten oder nur noch ein Formant wahrgenommen wird.

(nächste Folie)

Daher ist zu vermuten, dass bei gerundeten Vokalen (bei denen einer der

beiden eng beieinander liegenden Formanten abgeschwächt wird)tatsächlich dort nur eine Prominenz wahrgenommen wird.

(nächste Folie)

Wie wir wissen beeinflussen die Stimmqualität, aber auch das Zuschalten des Nasalraumes die spektrale Form des akustischen Outputs in den tieferen Frequenzbereichen.

Synthetisiert man ein Kontinuum von einem Vokal bis zu dem ihm entsprechenden Nasalvokal, so ist eine klare perzeptive Schwelle zu ermitteln. Interessant hieran ist, dass dies auch für Versuchspersonen zutrifft, deren Muttersprache keine Nasalvokale kennt.

Wird in Sprachsignalen die Amplitude der ersten Harmonischen modifiziert, kann wiederum eine klare Wahrnehmungsschwelle festgestellt werden. Wird dies mit männlichen Sprachsignalen getan, so entsteht eine Wahrnehmungsschwelle zwischen breathy und non-breathy. Zumindest gilt dies, wenn die Muttersprache der Versuchspersonen breathiness als distinktives Merkmal kennt.

Wird hingegen bei weiblichen Sprachsignalen die Amplitude der ersten Harmonischen erhöht, so ergibt sich die Wahrnehmungsschwelle nicht nasal - nasal. Vermutlich liegt dies an der Lage der ersten Harmonischen bei Frauenstimmen in der Nähe des ersten Formanten, wobei die Amplitudenerhöhung der ersten Harmonischen zu einer wahrgenommenen breiteren Bandbreite des ersten Formanten führen könnte, so dass dieser breiter als die Zwickerschen Frequenzbänder (Bark) werden könnte. Dies führt wie an anderer Stelle bereits erwähnt zum Eindruck der Nasalität.

(nächste Folie)

Eine weitere perzeptive Schwelle kann ermittelt werden, wenn einem tieffrequenten breitbandigen Signal ein Vokal, dessen Formanttransitionen entweder auf alveolare oder labiale Artikulationsstelle hinweisen, folgt. Ist die Amplitude des tieffrequenten Signals tiefer als eine bestimmte Stelle, so wird das Signal als stimmhafter Plosiv perzipiert, ist sie höher als die Schwelle, wird das Signal als Nasal interpretiert. Je nach scheinbarer Artikulationsstelle variiert diese Schwelle übrigens.

Somit korreliert dieses akustische Merkmal mit dem phonologischen Merkmal sonorancy. Das distinktive Merkmal continuincy korreliert dagegen mit der Präsenz oder der Abwesenheit eines abrupten Anstiegs der Amplitude in höheren Frequenzbereichen. (Frikativ-Affrikat Unterscheidung).

Wie wir wissen werden die Spektren von Konsonanten um so stärker von der Resonanz des Mundraumes verformt, je weiter hinten sie produziert werden. So

weisen Velare sogar formantähnliche Strukturen auf. Diese akustische Eigenschaften korrelieren mit dem Merkmal "compactness", der Phonologie, zumindestens der Lehre Jacobsons. Diese kompakten spektralen Strukturen werden vermutlich schmaler sein als die kritischen Bänder. Während hingegen das Rauschen alveolarer oder labialer Konsonanten mehrere kritische Bänder erreicht. Das heisst, Velare werden von labialen oder alveolaren unterschieden, da Velare eine einzelne spektrale Prominenz in mittleren Frequenzen aufweisen.

(nächste Folie)

Neben der spektralen Form kann auch die Amplitude des Gesamtsignals eines Konsonanten ausschlaggebend sein. So unterscheiden sich zum Beispiel /sa/ und /Sa/ dadurch, dass die Amplitude von /Sa/ höher ist als die des /s/ bei /sa/. Die Schwelle liegt hier interessanterweise so, dass sie der Amplitude des F3 von /a/ gleicht.

(nächste Folie)

Discussion:

Quantale Relationen sind als akustische, perzeptive und artikulatorische Korrelate

der distinktiven Merkmale der Phonologie anzusehen; für den Sprecher ist relativ wenig Präzision erforderlich.

Der Vorteil für den Hörer sind grosse Unterschiede in den akustischen/auditiven Mustern.

Es gilt FIGURE 1 auch in der Zeit: längere, stabilere Abschnitte stehen kurzen, starken Veränderungen gegenüber.

Grundsätzlich abstrahieren Quantalgebiete: so sind interindividuelle und intraindividuelle Unterschiede sowie Unterschiede zwischen Sprachen zwar vorhanden, aber sie behindern die Identifizierbarkeit sprachlicher Zeichen offenbar nicht.

Nicht alle Parameter sind gleichermassen stark quantal. Vermutlich werden die Quantaleren bevorzugt (Solche Präferenzen werden demnächst behandelt werden).

Es gibt auch häufig benutzte Features, die nicht offenkundig quantal sind, so z.B. Dauer.

Drei Fragen:

-Sind alle phonologisch-distinktiven Merkmale korrelierbar mit Quantalen Relationen?

-Welche akustischen Unterschiede haben das Potenzial, für phonetischen Kontrast benutzbar zu sein, und welche nicht (die aber dennoch im

Signal sind)?

-Welche artikulatorischen und akustischen Parameter bieten das grösste obengenannte Potenzial der Kontrastbildung? Wodurch können die Kontraste maximiert werden?