Spiel mit Lauten

Gerhard Jäger (Universität Bielefeld)

Wie viele Vokale gibt es im Deutschen? Das wären erst mal *a*, *e*, *i*, *o*, *u*, und dann noch die Umlaute *ä*, *ö* und *ü* – macht acht. Ein professioneller Phonetiker würde uns erklären, dass die Dinge noch ein bisschen schwieriger sind, aber nicht sehr. So könnte man noch kurze von langen Vokalen unterscheiden, schließlich klingt das *u* in *unten* ein bisschen anders als das in *Uhr*, nicht nur von der Länge her, sondern auch in der Vokalqualität. Außerdem machen wir zwar bei den Langvokalen einen Unterschied zwischen *e* und *ä* (etwa zwischen *Ehre* und *Ähre*), aber nicht bei den Kurzvokalen (*sengen* und *sängen* klingen gleich, trotz unterschiedlicher Orthographie). Dafür gibt es bei den Kurzvokalen noch den so genannten „Murmelvokal“ *Schwa*, der als *e* geschrieben wird und nur in unbetonten Silben vorkommt (wie das *e* in *lauf**e**n*). Insgesamt kommen wir also doch auf ein System von acht Kurzvokalen, plus weitere acht Langvokale.

Wer Fremdsprachen lernt, stellt fest, dass das nicht gottgegeben ist. Der ä-ähnliche Laut des Englischen, der zum Beispiel in *rat* vorkommt, ist nicht das selbe wie das Deutsche *ä*, und der Vokal im französischen *homme* ist zwar so ähnlich wie ein deutsches *o,* aber eben nur so ähnlich. Auch Laute aus verschiedenen Sprachen, die durch das selbe IPA-Symbol wiedergegeben werden, können sich subtil unterscheiden. So weist der niederländische Phonetiker Bart de Boer von der Universität Groningen darauf hin, dass das deutsche Wort *Kuh*, das niederländische *Koe* (‚Kuh'), das englische *coo* (‚girren') und das französische *cou* (‚Hals') durchaus verschieden klingen, obwohl sie alle durch *[ku:]* transkribiert werden.

Jede Sprache hat also ihr eigenes System von Vokalen. Wenn man sehr viele Sprachen vergleicht, stellt man aber fest, dass der Spielraum dabei dennoch recht begrenzt ist. So verfügen über 90% aller Sprachen über die Vokale *a, i* und *u*. Wenn eine Sprache nur drei Vokale hat (was selten ist, aber vorkommt), dann sind es so gut wie immer diese drei Vokale. Wenn eine Sprache fünf verschiedene Vokale hat, dann sind das so gut wie immer *a, e, i, o* und *u*. Die Liste derartiger allgemeingültiger, oder doch fast allgemeingültiger Gesetze ließe sich fortsetzen. Man kann sich also gleich über zwei Dinge wundern: warum unterscheiden sich die Vokalinventare verschiedener Sprachen so stark? Und warum unterscheiden sie sich nicht stärker?

Die Phonetiker (also Wissenschaftler, die sich mit der physiologischen und akustischen Realisierung von gesprochener Sprache befassen) haben in den letzten dreißig Jahren ein Verständnis dafür entwickelt, was ein „gutes“ Vokalsystem ist, und was es von einem schlechten unterscheidet. Bevor wir uns damit befassen können, müssen wir aber zunächst eine viel fundamentalere Frage klären: Was ist eigentlich ein Vokal?

Im Schulunterricht werden Vokale manchmal „Selbstlaute“ genannt, weil man sie alleine aussprechen kann, ohne Begleitlaut. Das ist aber eine unvollkommene Definition, weil das auch für die Konsonanten *l*, *r* oder *s* gilt. Eine genauere Definition nimmt darauf Bezug, wie Vokale artikulatorisch produziert werden:

- Vokale sind (wenn nicht gerade geflüstert wird) **stimmhaft**. Das bedeutet, dass die Stimmbänder (oder Stimmlippen, wie die Physiologen sagen), periodisch schwingen, was als Tonhöhe wahrgenommen wird.
- Der **Luftstrom** zwischen Lunge und Mund wird **nicht behindert**. Das unterscheidet Vokale von den Konsonanten. Bei letzteren ist der Luftstrom entweder zeitweise ganz unterbrochen (wie bei *b, p, k*), wird in die Nase umgelenkt (wie bei *m* und *n*), oder durch Zunge, Lippen oder Kehlkopf behindert.

Die einzelnen Vokale unterscheiden sich dabei durch die Position der Zunge und der Lippen. So ist zum Beispiel beim *u* die Zungenspitze relativ weit oben und hinten in der Mundhöhle, und die Lippen sind gerundet. Beim *i* ist die Zungenspitze oben vorne und die Lippen nicht gerundet usw. Ähnliche charakteristische Gesten der Artikulationsorgane lassen sich für jeden Vokal angeben.

Nun haben wir beim Hören selten die Gelegenheit, die Zungenspitze des Sprechers zu beobachten. Um Vokale (und Sprachlaute allgemein) zu identifizieren, sind wir auf die Schallwellen angewiesen, die der Sprecher produziert. Auch hier kann man jedem Vokal ein eindeutiges Muster zuweisen.

Akustisch gesehen entsprechen Vokalen periodische Schwingungen der Luftmoleküle (und, davon angeregt, des Trommelfells des Hörers). In Abbildung 1 ist ein Beispiel für die Amplitude des Vokals *u* dargestellt.

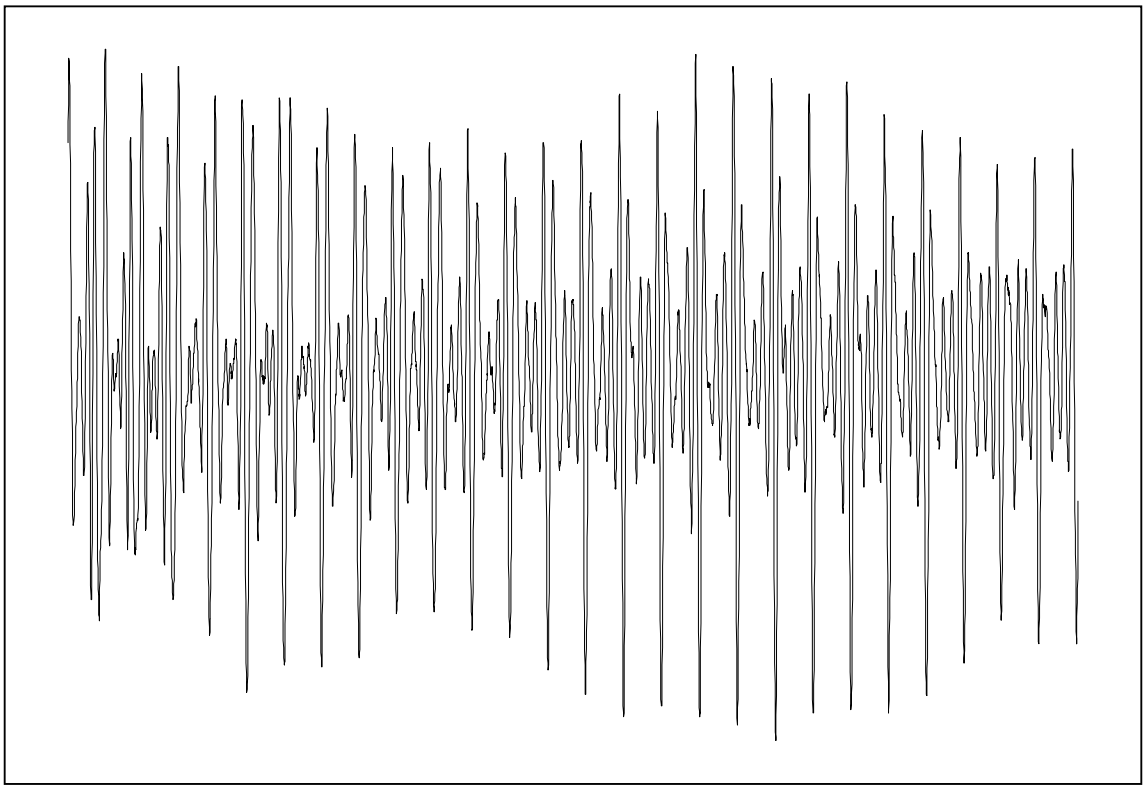


Abbildung 1: Amplitude, Vokal /u/

Nun enthält die Amplitude zwar alle nötige Information, aber selbst Experten benötigen mehrere Jahre Erfahrung, um in dieser Darstellungsform etwas ein /a/ von einem /o/ zu unterscheiden. Nützlicher ist da schon die Darstellung von Schallwellen in **Spektrogrammen**. Vereinfacht gesagt lässt sich jede periodische Schallwelle zerlegen in harmonische Schwingungen unterschiedlicher Frequenz, so wie sich weißes Licht durch die Spektralanalyse in Lichtwellen unterschiedlicher Frequenz zerlegen lässt. Ein Spektrogramm ist ein Diagramm, in dem die zeitliche Veränderung der Frequenzen dieser Komponenten dargestellt wird. Abbildung 2 zeigt ein Spektrogramm für die fünf Vokale *a*, *e*, *i*, *o*, *u*, gesprochen in dieser Reihenfolge, mit kurzen Pausen dazwischen.

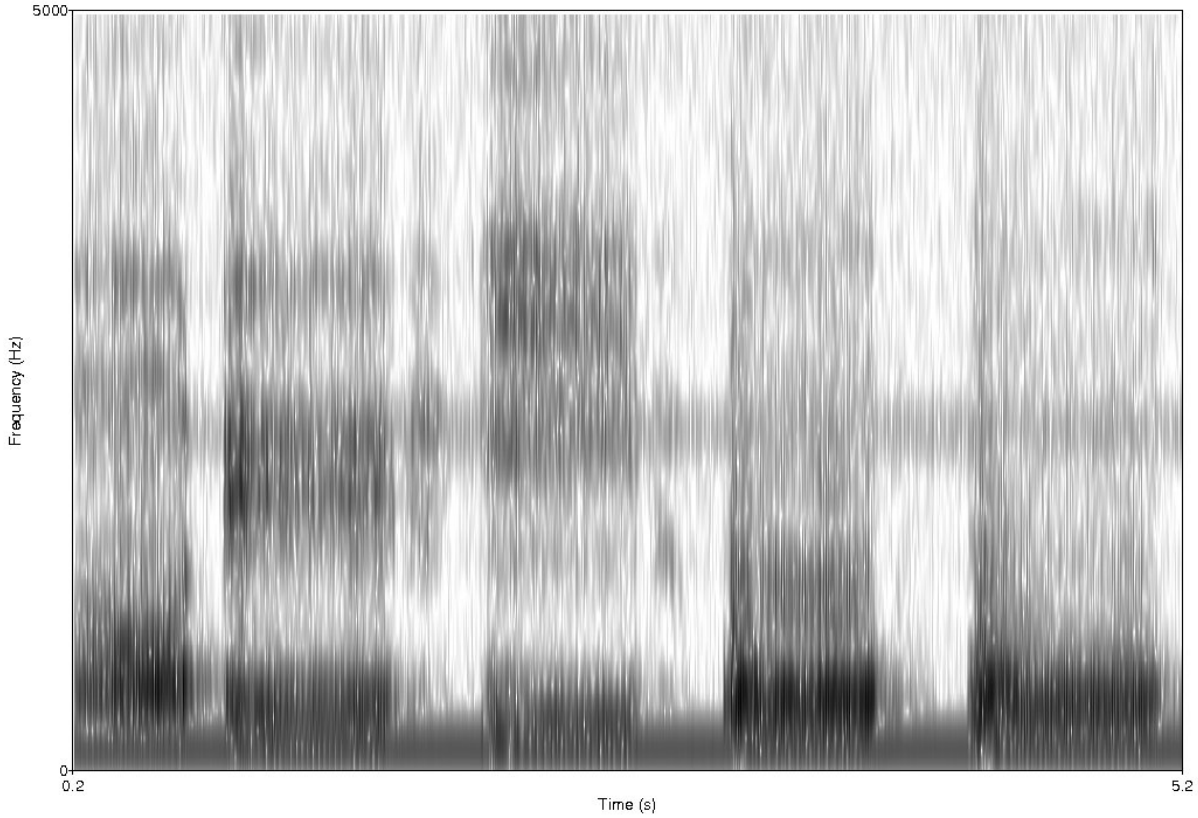


Abbildung 2 : Spektrogramm für die Vokalfolge „a – e – i – o – u“

Auf der senkrechte Achse ist die Frequenz dargestellt und auf der waagerechten die Zeit. Im untersten Bereich, bei ungefähr 75 Hz, ist ein durchgehender schwarzer Streifen zu sehen. Das ist die **Grundfrequenz**, mit der die Stimmlippen sich öffnen und schließen, und die wird als Tonhöhe wahrnehmen. Darüber sind mehrere verschieden starke dunkle waagerechte Bänder zu erkennen, die sich in der Tonhöhe von Vokal zu Vokal unterscheiden. In Abbildung 3 sind diese Frequenzbereiche speziell hervorgehoben. Hierbei handelt es sich um Formanten – Schwingungen, die der Grundfrequenz aufmoduliert sind. Ihre Frequenz hängt von den Resonanzeigenschaften der Mundhöhle und des Rachenraumes ab und kann durch die Position der Zunge, der Lippen usw. beeinflusst werden. Wie man in der Abbildung 3 deutlich erkennen kann, unterscheiden sich die Position der Formanten von Vokal zu Vokal.

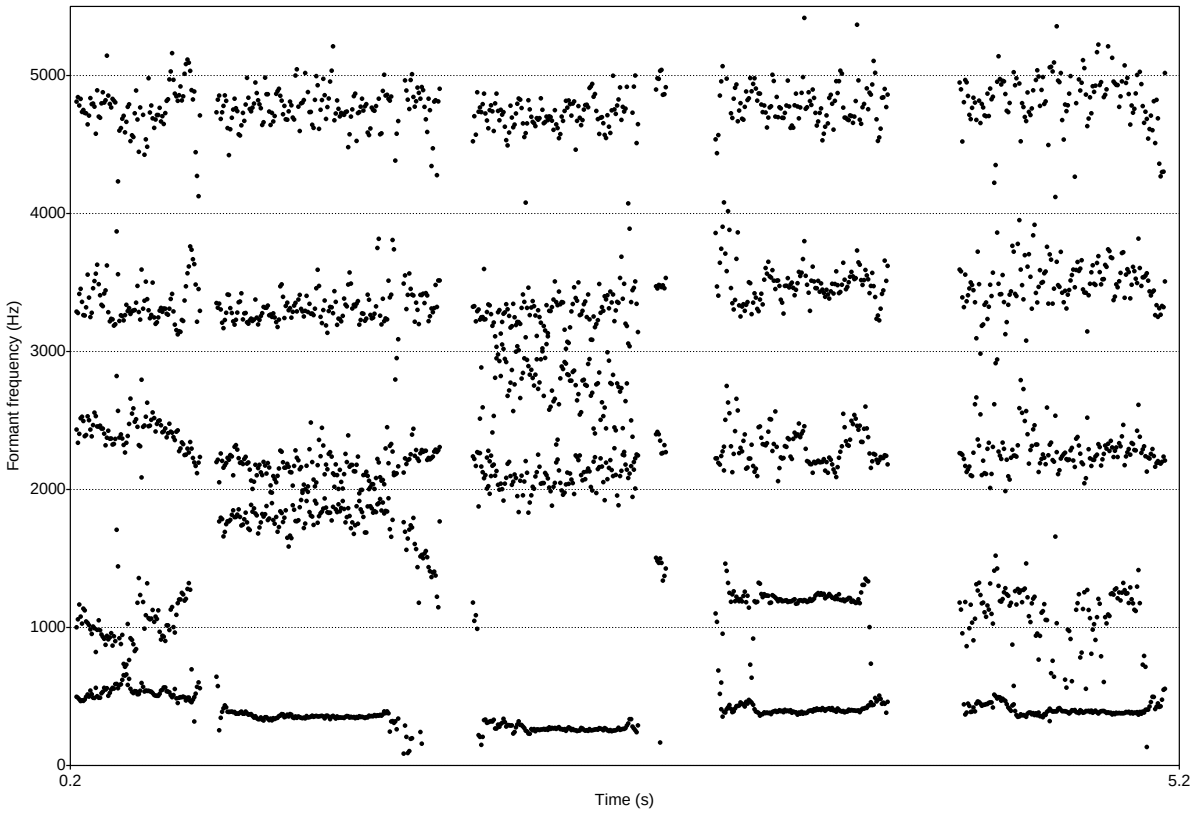


Abbildung 3: a-e-i-o-u: Formanten

Für den Vokalcharakter eines Sprachlautes sind ausschließlich die Formanten verantwortlich. Die Grundfrequenz spielt dafür keine Rolle. Es ist sogar so, dass die beiden tiefsten Formanten (also die beiden untersten im Spektrogramm) weitgehend bestimmen, als welcher Vokal ein Laut wahrgenommen wird. Der dritte Formant spielt noch eine geringe Rolle, während die höheren Frequenzbänder lediglich den individuellen Charakter einer Stimme ausmachen.

Man vereinfacht also nur wenig, wenn man sagt, dass eine bestimme Konfiguration von erstem und zweitem Formanten eine Vokalqualität definiert. Das ist praktisch für Zwecke der graphischen Darstellung. Man kann deshalb nämlich alle möglichen Vokalqualitäten in einem zweidimensionalem Diagramm darstellen.

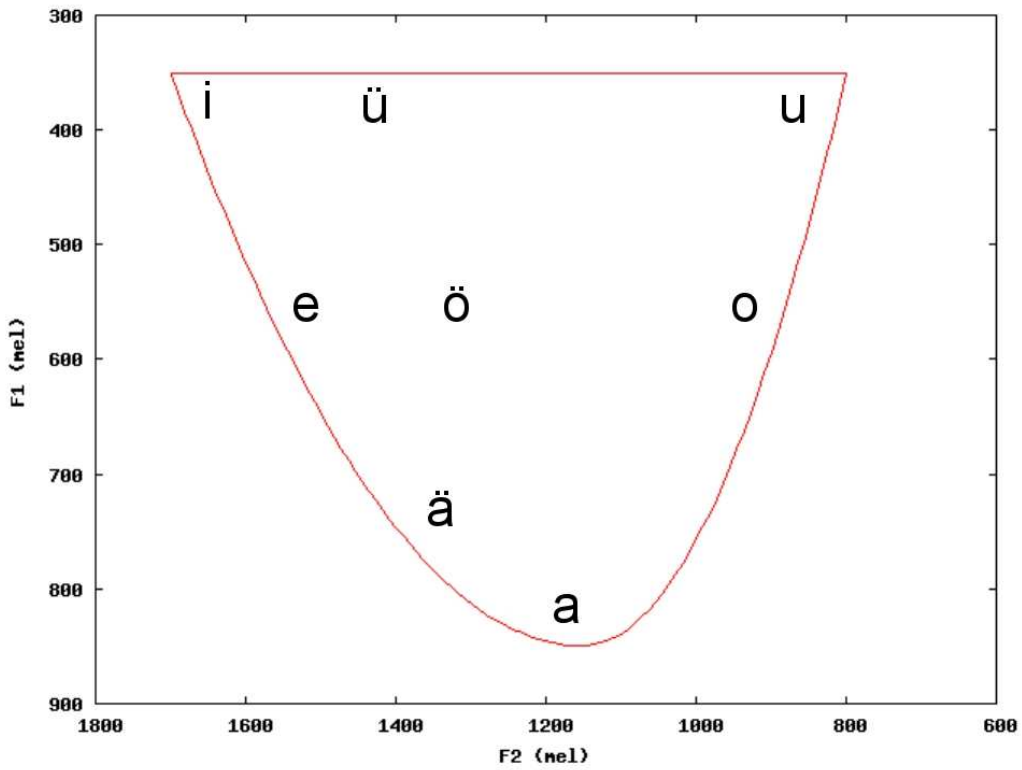


Abbildung 4: Idealisierte Positionen der deutschen Lang-Vokale im F1/F2-Raum

Abbildung 4 zeigt die Position der acht deutschen Langvokale in diesem zweidimensionalen Raum, der durch die beiden ersten Formanten gebildet wird. Auf der senkrechten Achse ist der erste Formant (F1) dargestellt und auf der waagerechten Achse der zweite Formant (F2). Statt der aus der Physik bekannten Maßeinheit Hertz werden Frequenzen in der Phonetik auch häufig in Mel dargestellt. Diese Maßeinheit bildet subjektiv empfundene Distanzen besser ab als Hertz.

Die Bereich der Vokalqualitäten, die mit der normalen menschlichen Anatomie produziert werden kann, ist in Abbildung 4 durch die rote Linie gekennzeichnet. Er hat eine paraboloide Form, die sich einem gleichschenkligen Dreieck annähert. *a*, *i* und *u* bilden die Ecken dieses Dreiecks.

Es ist wichtig zu erwähnen, dass die in Abbildung 4 angegebenen Positionen der deutschen Vokale Mittelwerte sind. Wenn man die tatsächlichen Realisierungen von verschiedenen Vorkommen eines Vokals im Laufe eines Monologes des selben Sprechers misst, erhält man eher ein Bild wie in Abbildung 5 – von Abweichungen zwischen verschiedenen Sprechern ganz zu schweigen.

Nachdem wir nun eine ungefähre Vorstellung davon gewonnen haben, was psychoakustisch gesehen ein Vokal ist, können wir uns wieder der ursprünglichen Frage zuwenden: Welche Vokalsystem gibt es in den Sprachen der Welt, und welche Systematik gibt es in der Variation zwischen den Sprachen? Um derartige Fragen beantworten zu können, wurde in den achtziger Jahren an der University of California in Los Angeles die „UCLA Phonological Segment Inventory Database“ aufgebaut. Unter Leitung des Phonetikers Ian Maddieson wurden das Lautinventar (also nicht nur Vokale, sondern auch Konsonanten) von mehreren hundert Sprachen kategorisiert. Die Auswahl der Sprachen geschah dabei so, dass alle existierenden

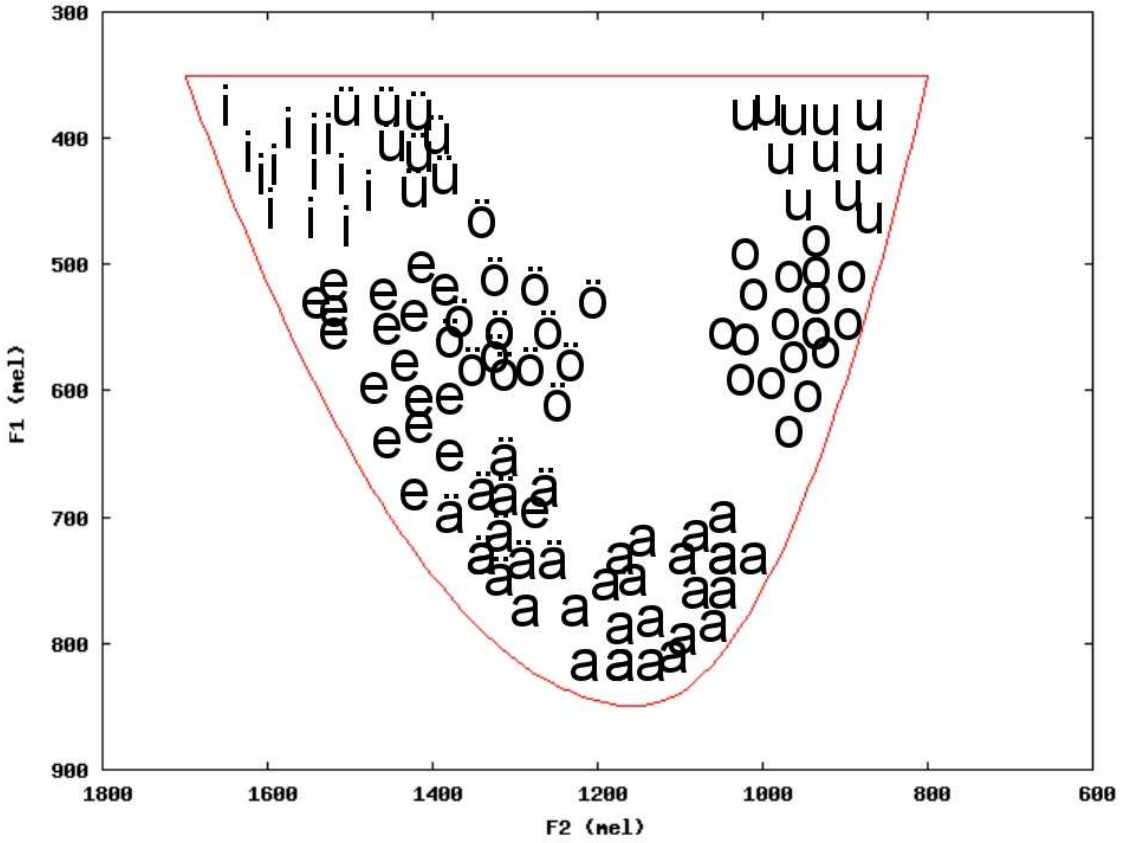


Abbildung 5: Die tatsächliche Realisierung von Vokalen weicht von den idealisierten Positionen mehr oder weniger stark ab.

Sprachfamilien möglichst gleichmäßig repräsentiert sind. Die französischen Forscher Jean-Luc Schwartz, Louis-Jean Boë, Nathalie Vallée und Christian Abry aus Grenoble benutzten diese Datenbank, um einen Katalog der häufigsten Vokalsysteme zu erstellen. Die wichtigsten Ergebnisse sind in der Graphik in Abbildung 6 dargestellt.

Fast alle Vokalsysteme umfassen zwischen 3 und 9 Vokalen (wobei z.B. für das Deutsche das System der Langvokale und das der Kurzvokale als zwei getrennte System betrachtet würden). Die Vokalsysteme sind zunächst geordnet nach der Anzahl der verschiedenen Vokale. Jede Zeile entspricht einer Anzahl von Vokalen. In jeder Zeile sind, von links nach rechts geordnet, die häufigsten Systeme mit dieser Anzahl von Vokalen schematisch dargestellt. Bei der Kategorisierung wurden nur sieben verschiedene Werte auf der vertikalen und maximal sechs Werte auf der horizontalen Achse unterschieden. Die Zahl links unten in jeder Zelle gibt an, wie häufig das entsprechende System in der Datenbank vertreten war.

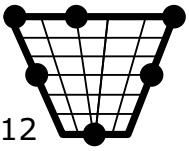
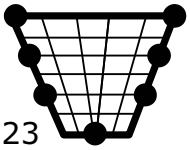
Anzahl der Vokale	Vokalsysteme und ihre Häufigkeit				
3	 14				
4	 14	 5	 4	 2	
5	 97	 3			
6	 26	 12	 12	4	
7	 23	 6	 5	 4	 3
8	 6	 3	 3	 2	
9	 7	 7	 3		

Abbildung 6: Die häufigsten Vokalsysteme – stilisierte Darstellung

Wenn man sich ein wenig in die Abbildung vertieft, erkennt man recht schnell mehrere auffällige Muster (die in der linguistischen Typologie schon seit der ersten Hälfte des zwanzigsten Jahrhunderts bekannt sind):

- Fast alle Vokale liegen an der Peripherie. Kein einziges System hat mehr als einen inneren Vokal. (Das Deutsche ist offensichtlich eine Ausnahmeerscheinung, da es bei den Kurzvokalen zwei innere Vokale gibt, das *ö* und den Murmelvokal Schwa.) Von den 264 Systemen, die erfasst sind, haben nur 64 überhaupt einen nicht-peripheren Vokal. In den meisten Fällen handelt es sich dabei um Schwa.
- Fast alle System haben ein *a*.
- So gut wie alle Systeme haben *i* und *u*. Die beiden einzigen Ausnahmen (in der zweiten Zeile) haben zumindest die ungespannten Varianten davon (wie das kurze *i* und das kurze *u* im Deutschen).

- Es gibt eine starke Tendenz zur Symmetrie. 215 der dargestellten 265 Systeme sind vertikal symmetrisch. (Vokale, die auf der dritten senkrechten Gitterlinie sitzen, sind hierbei als symmetrisch klassifiziert, da es bei sieben Gitterlinien nun einmal keine echte Mittelposition gibt.)

Schon 1972 hatten die schwedischen Forscher Johan Liljencrants und Björn Lindblom eine einfache, aber bestechende Idee, um diese Tendenzen zu erklären. Sie basiert auf der Einsicht, dass nahe nebeneinander liegende Vokale leicht zu verwechseln sind – sei es, weil der Sprecher ungenau artikuliert, sei es, weil die Übertragung des Schallsignals durch Nebengeräusche oder Ähnliches gestört ist. Ein ideales System sollte deshalb die Vokale im größtmöglichen Abstand voneinander anordnen, um die Verwechslungsgefahr zu minimieren. Die schwedischen Forscher verglichen dabei Vokale mit Magneten gleicher Ausrichtung, die auf einer Wasseroberfläche, etwa in einer Wasserschüssel, schwimmen. Wegen der wechselseitigen Abstoßung werden sie sich so weit wie möglich voneinander entfernen. Die Wasserschüssel hätte dabei die Form des Vokalraumes (vgl. Abbildung 4). Tatsächlich tendiert eine kleine Anzahl von Magneten, die auf Korkstückchen montiert sind, dazu, in einem entsprechend geformten Behälter die Positionen einzunehmen, die typologisch häufigen Vokalsystemen entspricht. Physikalisch gesprochen wären das die Zustände, in denen die potentielle Energie des Gesamtsystems von Magneten bzw. von Vokalen am Kleinsten ist.

In den vergangenen über dreißig Jahren wurde der Vorschlag von Liljencrants und Lindblom zwar in verschiedener Hinsicht modifiziert und verfeinert, aber viele Experten halten ihren Ansatz nach wie vor für fundamental korrekt: Optimale Vokalsysteme sind solche, in denen der Kontrast zwischen den Vokalen maximiert und damit die Verwechslungsgefahr in der Kommunikation minimiert wird. Daraus ergibt sich unmittelbar die Tendenz, Vokale an der Peripherie des Vokalraums anzuordnen. Die weite Verbreitung von *a*, *i* und *u* wiederum folgt in diesem Rahmen aus der besonderen Form des Vokalraums., eben weil seine Form näherungsweise ein Dreieck ist, und *a*, *i* und *u* die Ecken dieses Dreiecks bilden. Die Tendenz zur Symmetrie ergibt sich ebenfalls mehr oder weniger direkt aus der näherungsweisen Symmetrie des Vokalraums.

Natürliche Sprachen scheinen also clever organisiert zu sein – wenigstens was ihr Vokalinventar betrifft. Nun werden aber Sprachen nicht von Toningenieuren und Physiologen entworfen. Selbst Esperanto hat zwar eine sorgfältig konstruierte einfache Grammatik, aber es ist nicht bekannt, dass Ludovic Zamenhof (der Schöpfer von Esperanto) besondere Sorgfalt auf den Entwurf einer besonders durchkonstruierten Phonetik verwandte. (Esperanto folgt einfach der Mehrheit der Sprachen und hat die Vokale *a*, *e*, *i*, *o* und *u*.) Woran liegt es also, dass es zum Beispiel keine Sprachen mit dem Drei-Vokal-System *i, ü, e* gibt?

Ähnliche Fragen stellen sich in den verschiedensten Disziplinen, z.B. in der Biologie und in den Wirtschaftswissenschaften. Komplexe Systeme zeigen zuweilen eine verblüffende Zweckmäßigkeit, ohne dass sie für einen konkreten Zweck entworfen worden wären. So sind Pflanzen und Tiere scheinbar perfekt an ihre ökologische Nische angepasst. In einer Marktwirtschaft führt das freie Handeln der Marktteilnehmer dazu, dass Angebot und Nachfrage tendentiell in Übereinstimmung sind. Biologen erklären die Angepasstheit von Organismen an ihre Umwelt mit Hilfe der Evolutionstheorie. Wirtschaftswissenschaftler sprechen im Gefolge von Adam Smith von einer unsichtbaren Hand, die eine Volkswirtschaft als Ganzes ordnet. Das kausale Erklärungsmuster ist in beiden Fällen ähnlich: Wir haben es in jedem Falle

mit einer Population von Individuen zu tun – seien es Organismen einer Art oder Konkurrenten um eine Markt-Nische. Die einzelnen Individuen sind unterschiedlich gut an die Umgebung angepasst. Gut angepasste Organismen pflanzen sich erfolgreicher fort als die weniger glücklichen Artgenossen. Die Gene für gute Anpassung verbreiten sich deshalb, während unzweckmäßige Gene aussterben. Analog dazu animieren erfolgreiche wirtschaftliche Strategien zur Nachahmung, während weniger erfolgreiche Konzepte quasi aussterben. Das gemeinsame Muster ist, dass zweckmäßige Varianten (Gen-Varianten bzw. Verhaltensvarianten) eine größere Chance haben, wiederholt zu werden, sei es durch Nachahmung oder Fortpflanzung.

Die Analogie zwischen verschiedenen Erscheinungsformen von evolutionärer Selbstorganisation wurde von Biologen und Wirtschaftswissenschaftlern schon vor Jahrzehnten bemerkt. Unter dem Titel „Evolutionäre Spieltheorie“ gibt es sogar einen fachübergreifenden mathematischen Rahmen, in dem derartige Phänomene untersucht werden können. In den letzten Jahren greifen auch Vertreter anderer Geistes- und Sozialwissenschaften verstärkt auf evolutionäre Erklärungsmuster für kulturelle Phänomene zurück. So erregte der Philosph Daniel Dennett von der Tufts University bei Boston jüngst damit Aufmerksamkeit, dass er Religionen einer evolutionären Dynamik unterworfen sieht. Für den Erfolg einer Religionsgemeinschaft ist demnach primär nicht der Wahrheitsgehalt ihrer Lehre ausschlaggebend oder ihre Eignung als Opium des Volkes, sondern ihre Fähigkeit, sich in den Köpfen der Gläubigen „fortzupflanzen“. Dabei stehen verschiedene Religionen in Konkurrenz zueinander und sind einem Prozess der Auslese unterworfen.

Ein evolutionäres Erklärungsmuster ist auch auf die Organisation von Vokalsystem anwendbar (so wie auch auf viele andere Eigenschaften natürlicher Sprachen). Psycholinguistische Untersuchungen weisen darauf hin, dass wir unser phonetisches Wissen nicht in der Form von festen Regeln gespeichert haben, etwa der Art „Das *u* wird gebildet, indem die Zungenspitze nach oben hinten bewegt und die Lippen gerundet werden.“ oder „Wenn F1 bei 400 Mel liegt und F2 bei 900 Mel, handelt es sich um den Vokal *u*.“ Es ist eher so, dass wir ein gutes Gedächtnis haben und uns viele konkrete Beispiele von Vokalen, die wir in der Vergangenheit gehört oder selber produziert haben, schlicht merken. Wenn wir ein *u* sprechen wollen, imitieren wir einfach unsere eigenen Erfahrungen mit dem Versuch, ein *u* zu sprechen. Erfahrungen, die beim Hörer die Reaktion „Häh?“ auslösten, werden dabei eher ignoriert. Großhirn an Mund: Weißt du noch, was du das letzte Mal gemacht hast, als du ein *u* artikulieren solltest? Das war sehr gut – kannst du das noch mal machen? Oder: Dein letztes *u* war so grottenschlecht, mach das ja nicht noch mal! Analog bei der Sprachwahrnehmung: Ohr an Großhirn: Ich habe hier ein F1 von 392 Mel und F2 von 917 Mel – was soll ich'n damit machen? Großhirn an Ohr: Die letzten paar Mal, als wir so was Ähnliches hatten, haben wir es immer in den *u*-Topf gelegt. Bis jetzt hat sich niemand beschwert, also machen wir es diesmal wieder so!

Ein Sprecher wird also die Realisierungsarten eines Vokals am ehesten wiederholen, die am seltensten missverstanden werden. Das sind natürlich die, die im Vokalraum am Weitesten von konkurrierenden Vokalen entfernt sind. Um auf das Beispiel zurückzukommen: angenommen, eine Sprachgemeinschaft hat tatsächlich ein Vokalsystem, dass aus *i, e* und *ü* besteht. Die tatsächlich Realisierung der Vokale schwankt um den jeweiligen Mittelwert. Dabei werden Varianten des ü, die nahe beim Mittelwert des *i* liegen, leicht mit *i* verwechselt und deshalb nicht so häufig imitiert. Bei Varianten des *ü*, die mehr in Richtung *u* liegen, besteht (in dem genannten System) keine derartige Verwechslungsgefahr. Diese Varianten werden

eher imitiert. Deshalb verschiebt sich der Schwerpunkt der Realisierung des *ü* in Richtung *u*. Aus dem gleichen Grund verschiebt sich der Schwerpunkt der *e*'s in Richtung *a*. Ein stabiler Zustand wird erst erreicht, wenn aus dem *ü* ein *u* und aus dem *e* ein *a* geworden ist. Ein Vokalsystem *i, e, ü* ist also nicht grundsätzlich unmöglich, aber es würde sich binnen kurzer Zeit in Richtung *i, a, u* entwickeln.

Der bereits erwähnte Groninger Phonetiker Bart de Boer führte in den vergangenen Jahren umfangreiche Computersimulationen durch, bei denen künstliche Agenten über Vokale bzw. Formanten-Konfigurationen miteinander kommunizieren. Er konnte zeigen, dass eine quasi-evolutionäre Dynamik, wie sie im vorigen Abschnitt skizziert wurde, zwangsläufig zu „zweckmäßigen" Vokalsystem führt. Als zweckmäßig gelten natürlich solche Systeme, die den Kontrast zwischen den Vokalen maximieren – so wie die meisten natürlichen Vokalsysteme.

Eine stark vereinfachte Variante von de Boers Simulationen soll hier kurz vorgestellt werden. In der Simulationen gibt es 20 künstliche Agenten, die paarweise ein einfaches Spiel miteinander spielen. Zu Beginn des ganzen Spiels wird eine bestimmte Anzahl von „Vokalen" festgelegt und durchnummeriert. Am Anfang einer Runde wird per Zufall ein „Sprecher" und ein „Hörer" sowie die Nummer eines „Vokals" ausgewürfelt. Nur der Sprecher kennt die Identität des Vokals. Es ist seine Aufgabe, sie dem Hörer mitzuteilen. Dazu muss der Sprecher einen bestimmten Punkt innerhalb des Vokalraums (also ein F1/F2-Paar) auswählen und dem Hörer mitteilen. Der Hörer wiederum muss auf der Basis dieses Punktes im Vokalraum erraten, welcher Vokal gemeint war. Es kommen zwei erschwerende Bedingungen hinzu: Dem Sprecher zittert gewissermaßen die Hand bzw. die Zunge: er äußert nicht genau das F1/F2-Paar, auf das er abzielt, sondern der Punkt wird um einen zufälligen (normalverteilten) Betrag in eine zufällige Richtung verfälscht. Das Resultat dieser Verfälschung wird dem Sprecher nachträglich mitgeteilt. Auf diese Weise wir modelliert, dass Artikulation nicht völlig präzise ist. Der Hörer nimmt aber nicht diesen, schon verfälschten Punkt wahr. Stattdessen wird nochmals eine (ebenfalls normalverteilte) Zufallsvariable hinzugefügt. Das entspricht der Verfälschung eines akustischen Signals durch Nebengeräusche und ungenaue Wahrnehmung. Wenn der Hörer trotz dieser Hindernisse den korrekten Vokal errät, erhalten beide Spieler einen Punkt, andernfalls nicht.

Die Entscheidungskriterien der Spieler sind denkbar einfach. Jeder Spieler speichert vergangene Erfahrungen, sowohl Abbildungen von Vokalnummern auf F1/F2-Vektoren (im Sprechermodus) und als auch Kategorisierungen von F1/F2-Vektoren als Vokalnummern (im Hörermodus). In beiden Fällen wird einfach die Paarung Vokalkategorie–F1/F2-Vektor gespeichert, egal, ob sie im Sprecher- oder im Hörermodus benutzt wurde.

Zu Beginn des Spiels ist der Speicher eines jeden Spielers mit Zufallsabbildungen initialisiert. Wenn der Sprecher den Vokal v übermitteln soll, wiederholt er einfach eine erfolgreiche Artikulation von v aus seinem Speicher – bzw. er versucht sie zu wiederholen; wegen der unpräzisen Artikulation gelingt ihm das ja nur näherungsweise. Wenn der Hörer eine bestimmte Formantenkombination hört, durchsucht er sein Gedächtnis nach der erinnerten Beobachtung, die der aktuellen Wahrnehmung am ähnlichsten ist (die also im Vokalraum den geringsten Abstand zur aktuellen Beobachtung hat). Die aktuelle Beobachtung wird dann genauso kategorisiert wie dieser Gedächnisinhalt.

Wenn die Spieler am Ende einer Runde einen Punkt gewinnen, fügen sie die Erfahrungen aus dieser Runde ihrem Erfahrungsschatz hinzu (und vergessen dafür die älteste Erfahrung im Speicher, da der Speicher endlich ist). Wenn es keinen Punkt gab, wird die Erfahrung sofort vergessen.

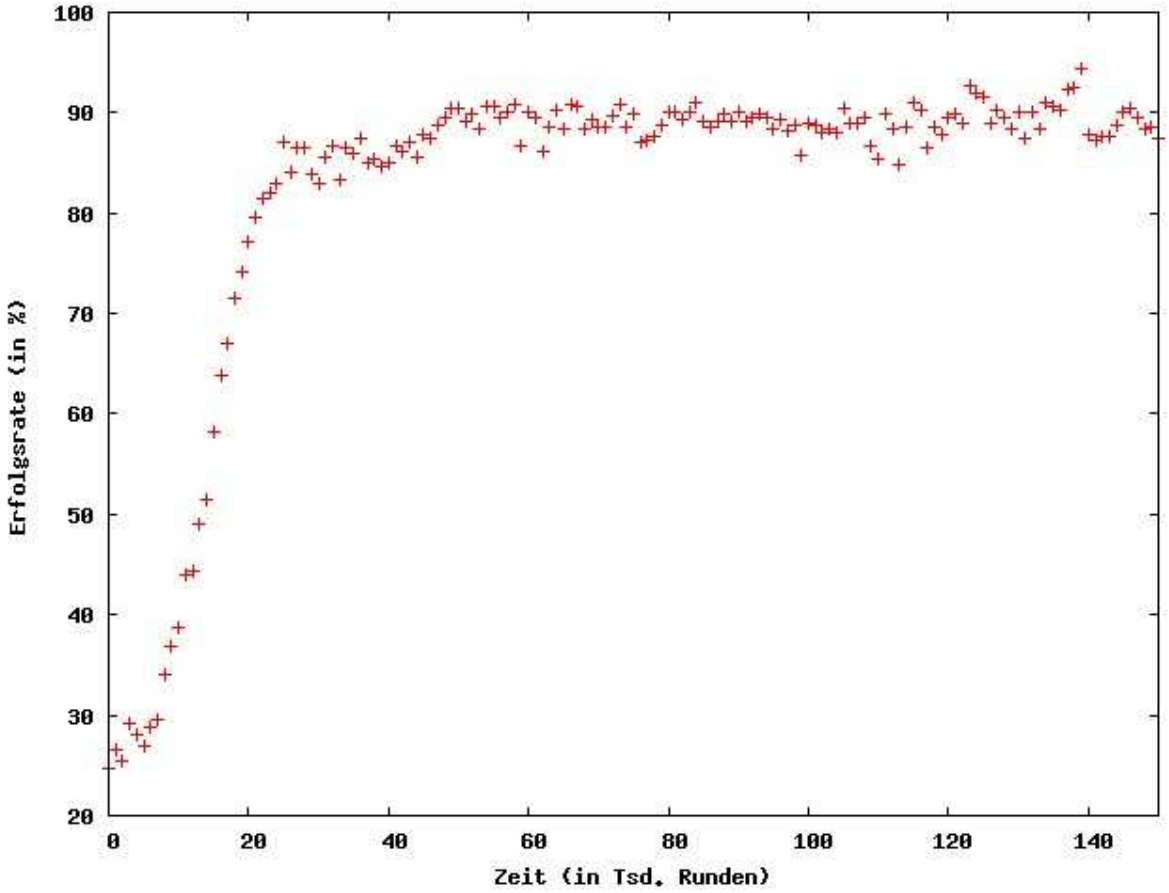


Abbildung 7: Während der Simulation bildet sich innerhalb der Population ein gemeinsamer Code heraus. Die Abbildung zeigt die Entwicklung der Erfolgsrate bei einem Spiel mit vier Vokal-Kategorien. Zu Beginn ist die Erfolgsrate 25% (= Zufallswert). Nach 50 000 Runden erreicht sie ein stabiles Niveau von 90%.

Ein Durchlauf einer Simulation lief immer nach dem selben Muster ab. Zu Beginn gibt es keinerlei Koordination zwischen Sprecher und Hörern. Deshalb haben sie allenfalls eine Zufallschance, einen Punkt zu gewinnen.

In der nächsten Phase bilden die einzelnen Agenten „private“ Kategorien. Damit ist gemeint, dass jeder Agent den einzelnen Vokalkategorien zusammenhängende Bereiche des Vokalraums zuordnet. Diese Bereiche können aber von Agent zu Agent stark voneinander abweichen. In einer weiteren Phase koordinieren sich die Kategorien der einzelnen Spieler. Dabei kann es vorübergehend Phasen geben, in denen in der Population verschiedene „Dialekte“ bestehen. Die Sprecher eines Dialektes benutzen zwar den selben Code, aber die Dialekte unterscheiden sich untereinander. Früher oder später bildet sich jedoch immer ein gemeinsamer Code heraus, der von der gesamten Population geteilt wird. Wenn dieser Zustand erreicht ist, bekommen die beiden Spieler fast in jeder Runde einen Punkt – sprich, Kommunikation funktioniert stabil. Abbildung 7 zeigt die Entwicklung der Erfolgsquote über einen längeren Zeitraum. Bei der Simulation dauerte es ungefähr 50 000 Runden, bis ein stabiles Plateau von 90% Erfolgsquote erreicht war.

In der letzten Phase verschieben sich die – nun von der gesamten Population geteilten – Kategorien innerhalb des Vokalraums so, dass der Abstand zwischen den

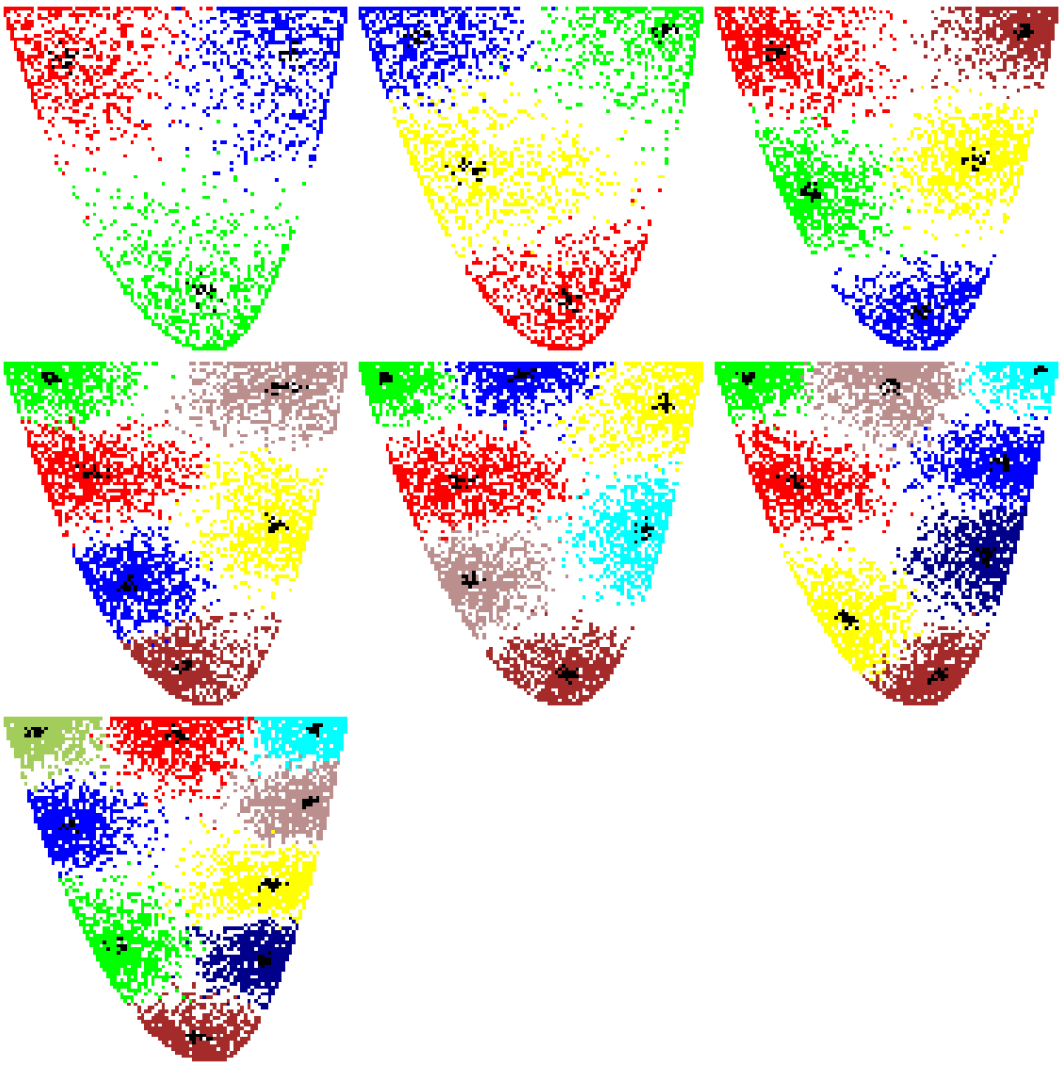


Abbildung 8: Simulations-Ergebnisse für 3-9 verschiedene vorgegebene Vokal-Kategorien. Die Graphiken geben den Gedächtnisinhalt der Agenten nach 300 000 Iterationen des Spiels an. Verschiedene Farben stehen für verschiedene Vokal-Kategorien. Die farbigen Punkte repräsentieren einzelne Ereignisse, die in wenigstens einem Gedächtnis gespeichert sind. Die schwarzen Punkte sind die Mittelwerte für die einzelnen Vokalkategorien für die einzelnen Agenten.

durchschnittlichen Realisierungen der einzelnen Kategorien maximiert wird. Wenn ein Zustand erreicht ist, in dem diese Werte maximal sind, befindet sich die Population in einem stabilen Gleichgewicht, dass sich auch über sehr lange Zeiträume nicht substantiell ändert. Allerdings kann es vorkommen, dass Zufallsfluktuationen sich aufaddieren und das System deshalb von einem stabilen Zustand (z.B. das Vier-Vokal-System *a, e, i, u*) in einen anderen stabilen Zustand (z.B. *a, o, i, u*) springt.

Abbildung 8 zeigt die stabilen Zustände, die für verschiedene vorgegebene Anzahlen von Vokal-Kategorien nach jeweils 300 000 Runden erreicht wurden. Die farbigen Quadrate stehen dabei für einzelne Sprech- oder Hör-Ereignisse, die am Ende des Spiels von wenigstens einem Agenten im Gedächtnis gespeichert waren. Die schwarzen Quadrate symbolisieren den Mittelwert der Realisierungen einer

Vokalkategorie, separat für jeden einzelnen Agenten. Wie leicht zu sehen ist, bilden die farbigen Punkte immer zusammenhängende Regionen. Innerhalb der Regionen liegen Cluster von schwarzen Punkten – es besteht also immer eine hohe Übereinstimmung zwischen den mittleren Gedächtnisinhalten der einzelnen Agenten.

Die stabilen Zustände in den Simulationen stimmen relativ gut mit den dominanten Tendenzen bei den natürlichen Vokalsystemen überein. Von den in Abbildung 6 dargestellten 264 Systemen entsprechen über die Hälfte, genauer gesagt 150, einem der in Abbildung 8 dargestellten stabilen Zustände. Andererseits sind von den sieben stabilen Zuständen aus Abbildung 8 fünf mehr oder weniger häufig in natürlichen Sprachen vertreten. Lediglich die Simulationsresultate für acht und für neun Vokal-Kategorien sind in Abbildung 6 nicht erfasst. Aber auch diese Systeme sind in gewisser Weise „natürlich“ - sie sind weitgehend symmetrisch, und die Schwerpunkte der emergenten Kategorien sind an der Peripherie des Vokalraums angeordnet. Das stärkste Defizit des Modells besteht darin, dass es nur periphere Vokale voraussagt, während der Zentralvokal Schwa in natürlichen Sprachen relativ häufig ist. Das hängt wohl mit der Tatsache zusammen, das Schwa leichter zu artikulieren ist als periphere Vokale – ein Aspekt, der in der Simulation nicht berücksichtigt wird. Die evolutionäre Dynamik, die durch häufige Iteration des oben geschilderten Spiels entsteht, liefert also eine plausible Erklärung dafür, warum Sprachen dazu tendieren, Vokale symmetrisch und an der Peripherie des Vokalraums anzuordnen.

Und was ist jetzt mit den Konsonanten? Auch bei Konsonanten-Inventaren gibt es klare universale Tendenzen. So haben zum Beispiel nahezu alle Sprachen Verschlusslaute (wie *p*, *t*, oder *b*) und Nasale (wie *m* oder *n*). Ein Übertragung des hier skizzierten Erklärungsansatzes auf Konsonanteninventare ist jedoch außerordentlich schwierig, und das aus mehreren Gründen. Anders als bei den Vokalen bildet der Raum der überhaupt artikulierbaren Konsonanten kein Kontinuum. So gibt es etwa keinen Laut, der irgendwo zwischen *p* und *t* angesiedelt ist. Auch ist es ziemlich schwierig, Konsonanten akustisch zu charakterisieren. Ein *k*, vor einem *a* klingt ganz anders als ein *k* vor einem *i* usw., auch wenn uns das nicht bewusst ist. Und von größeren sprachlichen Einheiten wie Silben, Wörtern und Sätzen haben wir noch gar nicht gesprochen. Es bleibt also spannend.

Weiterführende Literatur:

Bart de Boer, *The Origin of Vowel Systems*, Oxford University Press, 2001.

Daniel C. Dennett, *Breaking the Spell: Religion as a Natural Phenomenon*, Viking Adult, 2006.

Peter J. Richerson & Robert Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution*, University of Chicago Press, 2004.

John Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, 1982.

Pierre-Yves Oudeyer, *Self-Organization in the Evolution of Speech*, Oxford University Press, 2006.