

3 Sprachwahrnehmung

Barbara Höhle



Abbildung 4: Edison Phonograph *Excelsior* V201, Excelsior-Werke, Köln (um 1910)

Das Bild zeigt einen sogenannten Phonographen. Mit dieser von Thomas A. Edison entwickelten und im Jahre 1877 patentierten „Sprechmaschine“ war es zum ersten Mal möglich, Schall aufzuzeichnen und wiederzugeben. Der Trichter der Apparatur ist mit einer Membran versehen, die durch den einkommenden Schall in Schwingungen versetzt wird. Auf der Membran ist eine Nadel befestigt, die in Kontakt mit einer waagrecht angebrachten, drehbaren Walze steht, auf der eine Harzschicht aufgetragen ist. Die Nadel ritzt die beim Sprechen in den Trichter entstehenden Schwingungen der Membran in die Harzschicht ein. Dabei muss die Walze mit der Handkurbel gedreht werden. Zwar stehen heute ausgefeiltere Techniken zur Verfügung, aber die Möglichkeit, Sprachschall aufzunehmen, zu manipulieren und wiederzugeben, ist für die Sprachwahrnehmungsforschung weiterhin eine Grundlage ihrer Arbeit.

Die Forschung zur Sprachwahrnehmung beschäftigt sich mit den perzeptuellen und kognitiven Prozessen, die zwischen der Wahrnehmung des akustischen Signals durch den Hörer und seiner weiteren lexikalischen, syntaktischen und semantischen Verarbeitung liegen. Wie viele Wahrnehmungsprozesse läuft der Prozess der Sprachwahrnehmung weitgehend automatisch und unbewusst ab. Seine Effizienz erkennen wir daran, dass wir auch in einer lauten Umgebung meist noch gut in der Lage sind, sprachliche Äußerungen zu verstehen. Die Forschung zeigt jedoch, dass die Sprachwahrnehmung kein trivialer Prozess ist, sondern ein komplexes Verarbeitungssystem erfordert. Ein zentraler Untersuchungsgegenstand der Sprachwahrnehmungsforschung ist die Frage, auf welche Weise der Hörer in der Lage ist, in der Vielfalt und der Variabilität akustischer Informationen, die an sein Ohr dringen, diejenige phonologische Information zu erkennen, die er für weitere Sprachverarbeitungsprozesse nutzen kann. Speziellere Fragen hierbei sind: Welche akustischen Eigenschaften des Sprachsignals sind für die Erkennung von Sprachlauten wichtig? Welche Stufen der perzeptuellen Analysen werden durchlaufen? Welchen Effekt hat die Kenntnis einer bestimmten Sprache auf die Wahrnehmungsprozesse?

3.1 Das Problem der Sprachwahrnehmung

3.2 Die Wahrnehmung von Sprachlauten

3.3 Intermodale Sprachwahrnehmung

3.4 Theorien der Sprachwahrnehmung

3.1 Das Problem der Sprachwahrnehmung

Das sprachliche Signal stellt gleichzeitig das Produkt der menschlichen Artikulation und den Ausgangspunkt für den Wahrnehmungs- und Verständnisprozess gesprochener Sprache dar. Rein physikalisch betrachtet ist gesprochene Sprache ein akustisches Signal, das als Schallwelle das menschliche Ohr erreicht. Die Schallwelle ist durch bestimmte akustische Parameter wie Frequenz, Amplitude und deren Veränderungen über die Zeit gekennzeichnet. Die Frequenz des Signals hat einen Einfluss auf die wahrgenommene Tonhöhe, während die Amplitude die Wahrnehmung von Lautstärke bestimmt. Die beiden Parameter wirken dabei aber nicht unabhängig voneinander; so werden Töne mit höherer Frequenz grundsätzlich als leiser wahrgenommen als Töne derselben Amplitude mit niedrigerer Frequenz.

Um die akustischen Eigenschaften sprachlicher Signale zu verstehen, sind einige grundlegende Kenntnisse über deren Erzeugungsmechanismen während der Artikulation nötig. Nach der sogenannten Quellen-Filter-Theorie der Sprachproduktion unterscheidet man beim Sprechen das Quellsignal (ein von einer Schallquelle, z. B. den Stimmbändern erzeugtes Signal) und die Modifikation dieses Signals durch einen Filter. Bei stimmhaften Lauten stellen die Stimmbänder die Schallquelle dar, die durch den aus der Lunge kommenden Luftstrom in Schwingungen versetzt werden. Bei stimmlosen Lauten schwingen die Stimmbänder nicht, die Schallquellen sind in diesem Fall allein Verengungen im Vokaltrakt. Bei der Artikulation können gleichzeitig verschiedene Schallquellen beteiligt sein. Durch die Konstellation der Artikulatoren (im Wesentlichen Zunge, Lippen etc.) wird der in den Quellen erzeugte Schall auf unterschiedliche Weise modifiziert bzw. gefiltert. Dabei werden manche Frequenzbereiche des Schallsignals verstärkt, andere dagegen abgeschwächt. Dadurch entstehen Energiekonzentrationen in bestimmten Frequenzbereichen – die sogenannten Formanten. Die Formantenfrequenzen und deren Verhältnis zueinander sind ein wesentliches Merkmal für die Unterscheidung verschiedener Vokalqualitäten. Frikative (z. B. /f/, /s/) sind gekennzeichnet durch Turbulenzgeräusche, die durch Verengungen im Vokaltrakt (beispielsweise durch Anheben der Zunge) entstehen, sowie durch deren Frequenz, Dauer und Amplitude und ihre spektralen Übergangseigenschaften in die Umgebungsvokale. Ein typisches Merkmal von Plosiven (z. B. /p/, /d/) ist ein durch die Sprengung des Verschlusses entstehendes Geräusch (Burst). Das Frequenzspektrum

Quellen-Filter-
Theorie

Artikulatoren

Formanten

dieses Bursts wird bestimmt durch den Artikulationsort des Verschlusses. In Abhängigkeit vom Artikulationsort variieren aber auch die Formantenübergänge zu den umliegenden Vokalen. Diese Formantenübergänge sind spektrale Veränderungen des Signals, die aus Bewegungen der Artikulatoren resultieren.

Spektrogramm

→ **ABBILDUNG 5** zeigt das Spektrogramm für die drei Silben /pa/, /ba/ und /bi/. Ein Spektrogramm bildet die Intensität des Sprachsignals in verschiedenen Frequenzbereichen (y-Achse) über die Zeit (x-Achse) ab. Je höher der Grad der Schwärzung, desto höher ist die Intensität in diesem Frequenzbereich. Am Beginn des Signals sieht man bei allen drei Silben den Burst, der durch eine breite Verteilung über die Frequenzbereiche charakterisiert ist. Die im weiteren Verlauf des Signals erkennbaren dunklen Bänder bilden die Formanten ab, die gepunkteten Linien geben die mittleren Frequenzwerte dieser Formanten wieder. In diesem Beispiel erkennt man, dass die Formanten für die Silben /pa/ und /ba/ in den gleichen Frequenzbereichen liegen, während für die Silbe /bi/ der zweite Formant (von unten) in einem deutlich höheren Frequenzbereich liegt. Dies zeigt, dass die Formanten wesentliche akustische Eigenschaften der Vokale tragen. Vergleicht man die Silbe /pa/ mit /ba/ fällt auf, dass die klare Formantenstruktur bei der Silbe /ba/ schon früh zu Beginn des Signals erkennbar wird, während bei /pa/ eine deutliche Verzögerung im Einsatz der Formanten vor allem im unteren Frequenzbereich zu sehen ist. Diese Verzögerung wird als Stimmeinsatzzeit bezeichnet (→ **KAPITEL 3.2**), die bei stimmlosen Plosiven wie /p/ länger ist als bei stimmhaften wie /b/.

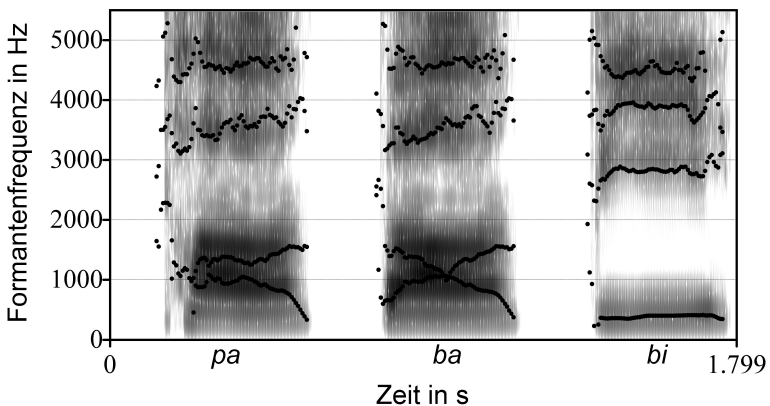


Abbildung 5: Spektrogramm für die Silben /pa/, /ba/ und /bi/

Einen der zentralen Untersuchungsgegenstände der Forschung zur Sprachwahrnehmung bildet die Frage, auf welche Weise der Hörer die Eigenschaften des Schallsignals nutzt, um das akustische Signal in sprachliche Information und somit in Bedeutung zu transformieren, d. h. anhand welcher Verarbeitungsmechanismen wird die akustische Information auf linguistische Einheiten wie Phoneme, Wörter, Phrasen und Sätze abgebildet. Mit den zunehmenden technischen Möglichkeiten, die akustischen Eigenschaften gesprochener Sprache nicht nur zu messen, sondern auch zu manipulieren, konnte die Forschung der Frage nachgehen, welche spezifischen akustischen Parameter im Signal für die Perzeption bestimmter segmentaler aber auch suprasegmentaler Eigenschaften relevant sind. Die entsprechenden Untersuchungen zeigten, dass das akustische Signal gesprochener Sprache durch zwei Eigenschaften charakterisiert ist, die dessen Dekodierung durch den Hörer zu einer komplexen Verarbeitungsaufgabe machen: mangelnde Invarianz und Nicht-Linearität.

Die mangelnde Invarianz ergibt sich daraus, dass es keine eindeutige Beziehung zwischen der Wahrnehmung eines bestimmten Lautes durch den Hörer und den akustischen Merkmalen des Signals zu geben scheint. Anders gesagt: bislang ist nicht geklärt, welche speziellen akustischen Hinweise im sprachlichen Signal dem Hörer die Identifizierung von bestimmten Sprachlauten ermöglichen. Für die Varianz der akustischen Merkmale der Realisierungen eines Lautes sind unterschiedliche Gründe ausschlaggebend. Durch die sogenannte Koartikulation zeigen sich Einflüsse der akustischen Eigenschaften, die benachbarte Lautsegmente aufeinander haben. Koartikulatorische Effekte sind ein Ergebnis der motorischen Steuerung beim Sprechen. Um eine flüssige Sprechweise zu erreichen, geht die Artikulation der einzelnen Laute ineinander über, d. h. die Artikulation erfolgt im Übergang von einem Segment zu anderen. Die genaue Konstellation der Artikulatoren bei der Produktion eines Lautes ist also immer mitbestimmt vom vorherigen und folgenden Segment. Beispielsweise sind die Formantenübergänge in Silben mit initialem Verschlusslaut (z. B. /bi/ vs. /ba/ vs. /bu/) nicht nur durch den Artikulationsort des Verschlusslautes bestimmt, sondern auch durch die Qualität des folgenden Vokals. Eine weitere Quelle der Variation ist die Position eines Lautes innerhalb einer Silbe oder eines Wortes. Plosive etwa werden in vielen Sprachen im Silbenanlaut aspiriert (angehaucht), nicht aber wenn sie im Auslaut einer Silbe auftreten. Auch die Betonung einer Silbe beeinflusst ihre akustische Realisierung. So weisen betonte Silben häufig eine höhere Dauer, eine höhere Intensität und eine hö-

**Mangelnde Invarianz
des Signals**

Koartikulation

here Grundfrequenz auf als unbetonte Silben. Nicht zuletzt tragen auch individuelle Merkmale des Sprechers zur akustischen Varianz bei. Jeder Sprecher verfügt über individuelle physiologische Merkmale, was die Größe und Konstellation des Artikulationstrakts und der Stimmbänder betrifft, die die Stimmhöhe und -qualität und auch die Realisierung einzelner Laute beeinflussen. Hierzu gehören systematische Unterschiede zwischen Männer-, Frauen- und Kinderstimmen, aber beispielsweise auch dialektale Färbungen. Auch affektive Zustände und andere Bedingungen der Sprechsituation beeinflussen die akustische Realisierung von Sprache. Diese individuellen und situationsbedingten Eigenschaften werden als indexikalische Eigenschaften des Sprachsignals bezeichnet.

→ **ABBILDUNG 6** zeigt, wie unterschiedlich verschiedene Sprecher die Vokale des Englischen realisieren. Gemessen wurde die mittlere Frequenz der unteren beiden Formanten (F1 und F2) der Vokale, die von 76 verschiedenen Sprechern produziert wurden. Die Ellipsen machen

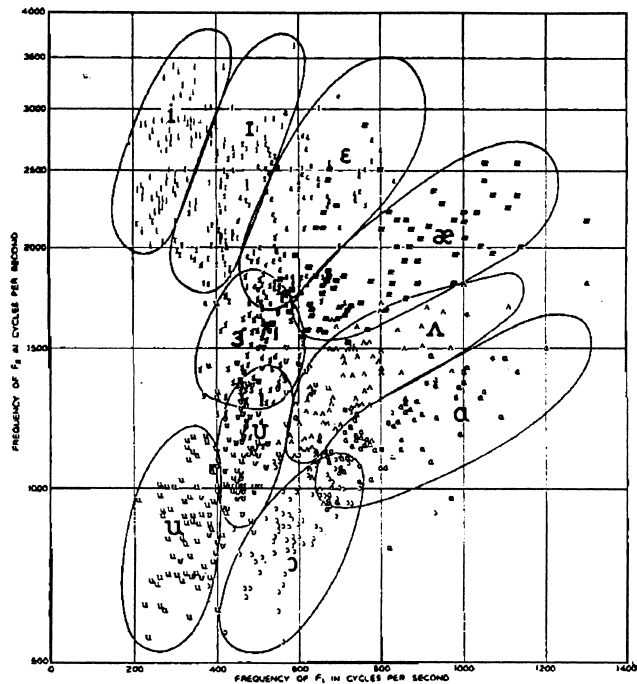


Abbildung 6: Mittlere Frequenzen des ersten Formanten (x-Achse) und des zweiten Formanten (y-Achse) verschiedener Vokale des Englischen

deutlich, wie Hörer diese Vokalrealisierungen kategorisieren, d. h. als welchen Laut sie die Produktion des Sprechers identifiziert haben. Neben der großen Streuung, die man innerhalb der Kategorien erkennen kann, ist zudem auffällig, dass es keine klaren Kategoriegrenzen gibt, sondern immer Mischungen der Kategorien.

Neben der mangelnden Invarianz ist die zweite wesentliche Eigenschaft des sprachlichen Signals seine Nicht-Linearität. Linearität würde erfordern, dass jedem Phonem (beispielsweise eines Wortes) genau ein Segment im akustischen Signal zugeordnet werden kann und dass die Abfolge dieser Segmente genau der Abfolge der Phoneme im Wort entspricht. Gesprochene Sprache ist jedoch kontinuierlich. Ihre Segmentierung in einzelne Laute ist meist nicht möglich, da sich im akustischen Signal die Eigenschaften benachbarter Phoneme vermengen. So sind beispielsweise die akustischen Hinweise, die eine Differenzierung zwischen verschiedenen Verschlusslauten (z. B. /b/ vs. /d/) erlauben, in den Formantenübergängen zum folgenden Vokal lokalisiert. Isoliert man das Geräusch der Sprengung des Verschlusses, so ist dieses Geräusch nicht als sprachlicher Laut erkennbar. Die Qualität des Lautes als stimmhafter Verschlusslaut mit bestimmtem Artikulationsort wird erst erkennbar, wenn der Vokal hörbar wird.

Nicht-Linearität
des Sprachsignals

3.2 Die Wahrnehmung von Sprachlauten

Trotz der erheblichen akustischen Variation in der Lautrealisierung, der mangelnden Invarianz und Nicht-Linearität sprachlicher Signale sind Hörer in der Lage, Laute zu identifizieren und sie Phonemkategorien zuzuordnen. Interessanterweise ist die Wahrnehmung lautlicher Unterschiede für verschiedene akustische Merkmale eine kategoriale. Kategoriale Wahrnehmung bedeutet, dass ein Hörer denselben akustischen Unterschied zwischen zwei Lauten dann besser wahrnehmen kann, wenn die beiden Laute unterschiedlichen phonologischen Kategorien (z. B. /b/ und /p/) zugeordnet werden als wenn sie derselben phonologischen Kategorie angehören (z. B. verschiedene Realisierungen von /b/).

Kategoriale
Wahrnehmung

Ein vor allem von einer Forschergruppe um den amerikanischen Phonetiker Alvin Liberman intensiv erforschtes Beispiel für die kategoriale Wahrnehmung stellt die Wahrnehmung der sogenannten Stimmeinsatzzeit (Voice-Onset-Time, VOT) dar (Liberman u. a. 1961). Als VOT bezeichnet man bei der Artikulation eines silbeninitialen Plosivs die zeitliche Relation, die zwischen dem bei der Spre-

Rolle der
Voice-Onset-Time

gung des Verschlusses entstehenden Geräusch und dem Einsatz der Stimmbandschwingungen für die Artikulation des folgenden Vokals besteht. Die VOT liefert also Information über die Stimmhaftigkeit des initialen Plosivs. Laute mit kurzer VOT (z. B. /ba/) werden als stimmhaft wahrgenommen, Laute mit längerer VOT (z. B. /pa/) als stimmlos. Für das Englische beispielsweise liegt die Grenze zwischen der Wahrnehmung eines stimmhaften und eines stimmlosen Plosivs in etwa bei 30–40 Millisekunden (ms). Durch die Manipulation der VOT kann man nun ein Kontinuum von Lauten erstellen, die sich lediglich in ihren Werten für die VOT (z. B. von 0 ms bis 80 ms) unterscheiden.

In Experimenten zur kategorialen Wahrnehmung präsentiert man Versuchspersonen beispielsweise ein auf diese Weise erzeugtes Lautkontinuum, bei dem sich die einzelnen Laute jeweils um 20 ms in ihrer VOT unterscheiden. Zur Überprüfung der kategorialen Wahrnehmung werden typischerweise Diskriminations- und Kategorisierungsaufgaben verwendet. Bei der Diskrimination (Unterscheidung) wird häufig das sogenannte AB–X Paradigma verwendet: Den Versuchspersonen werden drei Laute akustisch präsentiert, A und B sind Laute mit unterschiedlichen akustischen Eigenschaften, X entspricht entweder A oder B. Die Aufgabe besteht darin zu entscheiden, ob X mit A oder B identisch ist. Untersucht man mit diesem Paradigma die Wahrnehmung der VOT, würde man also in zufälliger Reihenfolge zum Beispiel folgende Lautkombinationen präsentieren: 0–20–0; 10–30–30; 20–40–20; 40–60–60, 60–50–60 usw. Bei der Kategorisierungsaufgabe wird den Versuchspersonen jeweils nur ein Laut (z. B. mit einer VOT von 0, 10, 20, ... bis 80) vorgegeben und sie sollen ihn identifizieren, indem sie den Laut benennen.

Kategoriale Wahrnehmung zeigt sich in diesen Aufgaben an folgenden Mustern: In der Diskriminationsaufgabe ergeben sich unter- und oberhalb einer bestimmten VOT-Grenze Leistungen auf Rate-niveau, d. h. die Versuchspersonen entscheiden bei einer gegebenen Lautfolge AB–X in 50 % der Fälle, dass X A entspricht und in 50 % der Fälle, dass X B entspricht. Bei Lautsequenzen, deren VOT über die kritische Grenze hinweggeht, zeigen sich dagegen sehr gute (im Idealfall 100 % übereinstimmende) Zuordnungen. Ein Beispiel: Bei den Lautfolgen 0–20–0 und 10–30–30 erzielen die Probanden Zufallsleistungen, bei der Lautfolge 20–40–20 lässt sich dagegen feststellen, dass mit nahezu 100 %iger Übereinstimmung X A zugeordnet wird; bei 40–60–60 und 70–50–70 zeigen die Probanden dagegen wieder Leistungen auf Zufallsniveau. Ein solches Ergebnis

Diskrimination und
Kategorisierung

Evidenz für
kategoriale
Wahrnehmung

deutet darauf hin, dass Laute mit einer VOT von kleiner 30 bzw. größer 40 untereinander nicht diskriminiert werden können, Laute mit einer VOT von 20 aber sehr gut von Lauten mit einer VOT von 40 diskriminiert werden können. Daraus lässt sich auf eine Kategoriegrenze zwischen der Wahrnehmung eines stimmhaften und eines stimmlosen Plosivs schließen, die zwischen einer VOT von 20 und 40 ms liegt. Bei der Kategorisierungsaufgabe würde man entsprechend erwarten, dass etwa alle Laute bis zu einer VOT von 30 ms eindeutig als /ba/ identifiziert und alle Laute mit einer VOT ab 40 ms eindeutig als /pa/ identifiziert werden.

Wahrnehmung von
Kategoriegrenzen

Befunde dieser Art gaben Anlass zu der Vermutung, dass die menschliche Sprachwahrnehmung auf phonologisch relevante Eigenschaften des sprachlichen Signals eingestellt ist und akustische Information, die nicht phonemunterscheidend wirkt, ignoriert. Grundsätzlich scheint die kategoriale Wahrnehmung eine generelle Eigenschaft des akustischen Systems zu sein, sie findet sich bereits bei Neugeborenen (Eimas u. a. 1971), wurde bei einigen Tierarten nachgewiesen (Kuhl/Miller 1975) und ist beim Menschen nicht auf die Wahrnehmung von Sprache beschränkt (Pisoni 1977). Untersuchungen mit feineren Analyseinstrumenten weisen jedoch darauf hin, dass die kategoriale Wahrnehmung nicht so stark ist, wie sich in den ersten Diskriminations- und Kategorisierungsaufgaben zeigte. Eine Messung von Reaktionszeiten ergab, dass diese verlängert waren, wenn die zu beurteilenden Stimuli sich in der Nähe der Kategoriegrenzen befanden (Pisoni/Tash 1974). Zudem können Hörer Unterschiede in der Güte eines Lautes wahrnehmen. Typischerweise werden Laute, deren akustische Eigenschaften nahe an der Kategoriegrenze liegen, als weniger gute Exemplare der Kategorie beurteilt (Miller 1997). Das bedeutet, dass die nicht phonemunterscheidende Information sehr wohl wahrgenommen und keineswegs ignoriert wird.

Kategoriale
Wahrnehmung
als Eigenschaft des
akustischen Systems

Ein anderer Effekt wurde bei der Wahrnehmung von Vokalen beobachtet. Vokale werden weniger stark kategorial wahrgenommen als Konsonanten, Hörer verfügen hier auch über bessere innerkategoriale Unterscheidungsfähigkeiten (Fry u. a. 1962). Diese Unterscheidungsfähigkeit wird allerdings von der Prototypikalität beeinflusst, die ein Vokalexemplar für seine Kategorie besitzt. Patricia Kuhl fand heraus, dass Sprecher derselben Sprache übereinstimmend darüber urteilen, welche akustischen Exemplare einer Vokalkategorie besonders gute bzw. typische Repräsentanten (Prototypen) und welche weniger gute bzw. atypische Repräsentanten dieser Kategorie darstellen. Experimente mit als gut und als weniger gut eingestuften Exemplaren

Perzeptueller
Magneteffekt

von Vokalen zeigten, dass es Sprechern der Sprache schwerer fiel, ein weniger typisches Exemplar der Vokalkategorie von einem typischen zu unterscheiden, wohingegen sie zwei wenig typische Exemplare derselben Kategorie besser voneinander unterscheiden können. Dieser Effekt wird als perzeptueller Magneteffekt bezeichnet (Kuhl 1991). Diese Befunde zeigen, dass Vokalkategorien eine inhärente Struktur aufweisen, die die Wahrnehmung von Exemplaren der Kategorie beeinflusst.

Der Einfluss
der Sprache auf die
kategoriale
Wahrnehmung

Weitere Untersuchungen zeigen, dass unsere Sprachwahrnehmung von unserer linguistischen Erfahrung beeinflusst wird. Arthur Abramson und Leigh Lisker fanden in einer Diskriminationsaufgabe heraus, dass dasselbe lautliche Kontinuum von Sprechern verschiedener Sprachen unterschiedlich kategorisiert wird. Während englische und spanische Sprecher auf einem Lautkontinuum jeweils zwei Kategorien unterschieden, die aber sprachspezifisch unterschiedliche Kategoriegrenzen aufwiesen, unterschieden Sprecher des Thai auf demselben Lautkontinuum drei Kategorien. Dies wird darauf zurückgeführt, dass im Thai im Gegensatz zum Englischen und Spanischen aspirierte Plosive eine eigene phonologische Kategorie darstellen (Abramson/Lisker 1970).

Wahrnehmung nicht
muttersprachlicher
Laute

Der Einfluss des muttersprachlichen phonologischen Systems zeigt sich besonders eindrucksvoll bei der Wahrnehmung von Lautkontinua aus einer fremden Sprache, die in der eigenen Muttersprache keinen phonologischen Status haben. Beispielsweise fanden Janet Werker und Richard Tees, dass erwachsene englische Muttersprachler den Kontrast zwischen einem dentalen und einem retroflexen Laut des Hindi nur schlecht unterscheiden konnten, während dies erwachsenen Sprechern des Hindi mühelos gelang (Werker/Tees 1984). Der Einfluss der sprachlichen Erfahrung auf die Wahrnehmung beginnt sich bereits in der frühesten Kindheit zu etablieren. In ihrer Studie zum dental-retroflex Kontrast im Hindi untersuchten Werker und Tees auch englische Kinder zwischen dem 6. und 12. Lebensmonat. Während die Säuglinge im 6. und 8. Lebensmonat diese Kontraste noch gut unterscheiden konnten, hatte die Unterscheidungsfähigkeit mit 12 Monaten bereits deutlich nachgelassen und war nicht mehr besser als bei den englischen Erwachsenen. Derselbe Verlauf zeigte sich bei der Untersuchung des im Japanischen nicht existierenden /r/-/l/ Kontrastes bei japanischen Säuglingen: 6-monatige konnten diese Laute noch gut unterscheiden, 12 Monate alte zeigten bereits die gleichen schlechten Leistungen wie Erwachsene. Bei ausreichend spezifischem Training können aber auch Erwachsene

noch lernen, nicht-native Kontraste zu differenzieren (Logan u. a. 1991).

Auch das Auftreten des perzeptuellen Magneteffekts ist ein Ergebnis der sprachlichen Erfahrung. Er zeigt sich ausschließlich bei der Wahrnehmung muttersprachlicher Vokale, bei nicht-nativen Vokalen tritt er nicht auf. Diese Beobachtung spricht ebenso für eine Abhängigkeit des perzeptuellen Magneteffekts von Spracherfahrung wie die Tatsache, dass der Effekt weder bei nicht-humanen Spezies noch bei Säuglingen auftritt, die jünger als 6 Monate sind (Kuhl 1991).

3.3 Intermodale Sprachwahrnehmung

Die Forschung hat sich auch intensiv mit der Frage beschäftigt, inwiefern die Sprachwahrnehmung von anderen Wahrnehmungsquellen profitiert, vor allem von der visuellen Information, die durch die Bewegungen des Gesichts des Sprechers während der Artikulation gegeben ist.

**Einfluss visueller
Information auf die
Sprachwahrnehmung**

Eines der bekanntesten Phänomene, das den Einfluss visueller Information auf die Lauterkennung demonstriert, ist der sogenannte McGurk-Effekt. In einer vielbeachteten Studie aus dem Jahr 1976 präsentierten die Wahrnehmungsforscher Harry McGurk und John MacDonald ihren Probanden nicht-kohärente visuelle und auditive Information: sie zeigten ihnen beispielsweise ein Gesicht, das die Silbe /ga/ artikulierte, und spielten ihnen gleichzeitig auditiv die Silbe /ba/ vor. Die Aufgabe der Probanden bestand darin, die auditiv präsentierte Silbe zu identifizieren. Es zeigte sich, dass die Mehrzahl der Probanden antwortete, sie hätten die Silbe /da/ gehört – also eine Silbe mit einem initialen Konsonanten, dessen Artikulationsort *zwischen* denen der initialen Konsonanten der gehörten und der „gesehenen“ Silben liegt (McGurk/MacDonald 1976).

McGurk-Effekt

Dass der auditive Kanal tatsächlich nicht die einzige Informationsquelle für die Sprachwahrnehmung darstellt, zeigt sich auch daran, dass diese erleichtert wird, wenn der Hörer während der Artikulation das Gesicht des Sprechers sehen kann (Massaro/Jesse 2007). Die gleichzeitige visuelle Information wirkt erleichternd, sowohl bei der Identifizierung einzelner Laute als auch bei der Verarbeitung von Wörtern, ganzen Sätzen und ganzen Textpassagen, sowohl bei klarer Sprache als auch in Bedingungen, in denen das Sprachsignal durch Störgeräusche überlagert wird. Offensichtlich ergänzen sich die auditive und visuelle Wahrnehmung bei der intermodalen Sprachwahr-

Ergänzung von visueller und akustischer Information

nehmung (d. h. wenn an der Wahrnehmung mehrere Sinnesmodalitäten beteiligt sind) in optimaler Weise. Die Wahrnehmung von Artikulationsart und Stimmhaftigkeit wird nur wenig von gleichzeitiger visueller Information beeinflusst, sie ist allerdings in der auditiven Wahrnehmung auch unter Störbedingungen besonders robust. Die Wahrnehmung des Artikulationsortes eines Lautes wird dagegen stärker von visueller Information beeinflusst, da dieser in einem Gesicht offenbar einfacher erkennbar ist. Auch hier findet sich ein feines Zusammenspiel zwischen akustischer und visueller Information. Während die hinteren Artikulationsorte im akustischen Signal relativ gut kodiert und daher gut auditiv erkennbar sind, gilt dies nicht für Labiale, deren vordere Bildungsstelle (an den Lippen) jedoch der visuellen Wahrnehmung besonders gut zugänglich ist. Neben der Lippenposition tragen auch die Zungenposition und die Sichtbarkeit der Zähne zu der unterstützenden Wirkung visueller Information bei der Sprachlauterkennung bei.

3.4 Theorien der Sprachwahrnehmung

Motortheorie der Sprachwahrnehmung

Theorien der Sprachwahrnehmung haben sich der Frage gestellt, wie es dem Menschen gelingen kann, trotz der mangelnden Invarianz und der Nicht-Linearität des akustischen Signals gesprochene Sprache zu dekodieren. Eine einflussreiche Theorie, die zur Lösung des Varianzproblems bei der Lautwahrnehmung vorgeschlagen wurde, stellt die von Alvin Liberman Mitte der 1980er-Jahre propagierte Motortheorie der Sprachwahrnehmung dar (Liberman/Mattingly 1985). Nach dieser Theorie sind die Programme für die Artikulationsgesten, die die Produktion von Lauten motorisch steuern, eine zentrale Instanz für die Wahrnehmung und Identifizierung von Lautsegmenten. Diese Programme werden als invariant verstanden, die akustische Varianz der Lautsegmente entsteht erst bei der Ausführung der Artikulationsgesten. Beim Hören gesprochener Sprache versucht der Hörer demnach, die Artikulationsgesten zu identifizieren, mit dem die gehörten Laute produziert wurden. Darüber hinaus nimmt die Motortheorie an, dass die Wahrnehmung von Sprache aufgrund der spezifischen Eigenschaften von Sprache ein sprachspezifisches phonetisches Wahrnehmungssystem erfordert, das als Ergänzung zum generellen auditiven System operiert.

Theorie des direkten Realismus

Eine enge Kopplung von produktiven und perzeptuellen Mechanismen nimmt auch die Theorie des direkten Realismus an (Fowler

1986). Im Gegensatz zur Motortheorie wird hier als Gegenstand der Wahrnehmung nicht die intendierte Artikulationsgeste, sondern die Artikulationsgeste selbst betrachtet. Dieser Ansatz wird vor allem durch den erleichternden Effekt visueller Information auf die Sprachwahrnehmung unterstützt. Zudem nimmt die Theorie des direkten Realismus nicht an, dass es ein sprachspezifisches Modul der Wahrnehmung gibt, sondern geht davon aus, dass die Sprachwahrnehmung allein anhand genereller domänen-unspezifischer auditiver Mechanismen erfolgt.

Beide Theorien machen wenig konkrete Annahmen dazu, wie die Kopplung des auditiven Ereignisses an die Artikulationsmotorik erfolgt. Neuere Stützen für motorisch basierte Ansätze bietet aber die Entdeckung der sogenannten Spiegelneuronen (Rizzolatti/Craighero 2004), die neuronale Aktivität sowohl zeigen, wenn ein Organismus eine bestimmte Handlung durchführt, als auch wenn diese Handlung nur beobachtet oder ein typisches mit dieser Handlung verbundenes Geräusch gehört wird. Da der Sprecher bei der Artikulation das von ihm selbst erzeugte sprachliche Signal auch hört, könnte sich eine sehr enge Assoziation zwischen motorischen und akustischen Mustern bilden.

Spiegelneuronen

Andere Theorien gehen dagegen von einer Dominanz akustischer Information bei der Sprachwahrnehmung aus und nehmen an, dass invariante akustische Merkmale im akustischen Signal existieren, die die Identifizierung von Lauten ermöglichen. Anhand einer differenzierteren Analyseweise fanden Forscher Hinweise auf invariante akustische Merkmale für den Artikulationsort eines Lautes, wenn man die gegebene spektrale Information in sehr kleinen Zeitfenstern von 10–25 ms analysiert. Dabei zeigten sich innerhalb der ersten 25 ms einer Konsonant-Vokal-Silbe mit initialem Plosiv Energiekonzentrationen im niedrigen Frequenzbereich für Bilabiale, im mittleren Frequenzbereich für Velare und im höheren Frequenzbereich für Alveolare. Auf diesen Befunden basiert die Theorie der Merkmalsdetektoren, die genau diese invarianten spektralen Eigenschaften des Signals detektieren und somit speziell auf die Eigenschaften des Signals reagieren, die für ihre Identifikation notwendig sind (Stevens/Blumstein 1980).

Theorie der Merkmalsdetektoren

Das Problem der Invarianz umgehen exemplar-basierte Modelle, die davon ausgehen, dass eine perzeptuelle Kategorie durch alle bislang gehörten Exemplare dieser Kategorie mit all ihren akustischen Eigenschaften repräsentiert wird. Jedes Exemplar besteht aus einer Assoziation zwischen einer Menge akustischer Eigenschaften, d. h.

Exemplar-basierte Modelle

detaillierter spektraler Information, und einer Menge von Kategoriela-beln. Jedem Exemplar sind daher auch linguistische sowie indexika-lische Informationen z. B. über das Geschlecht des Sprechers zugeord-net. Bei der Identifizierung neuer sprachlicher Signale kann jegliche hier gespeicherte Information herangezogen werden. Diese Modelltypen können daher auf einen Normalisierungsprozess verzichten, der die Va-rianz aus dem Sprachsignal herausfiltert. Tatsächlich zeigt empirische Evidenz vor allem aus dem Bereich der Wortverarbeitung, dass indexi-kalische Eigenschaften des Signals gespeichert werden und bei der Worterkennung einen Einfluss haben (Goldinger 1998). Bislang ist aber unklar, ob indexikalische und linguistische Information gemeinsam re-präsentiert werden oder ob sie separate Speicher- und Verarbeitungs-mechanismen ansprechen.

Fragen und Anregungen

- Was bedeuten die Begriffe der Varianz (bzw. mangelnden Invari-anz) und der Nicht-Linearität des sprachlichen Signals?
- Überlegen Sie, vor welches Problem die Eigenschaften der man-gelnden Invarianz und der Nicht-Linearität die Sprachwahrneh-mungsforschung stellen.
- Erläutern Sie, mit welchen Aufgaben die kategoriale Wahrneh-mung überprüft wird.
- Diskutieren Sie, ob die sprachliche Erfahrung Einfluss auf die Sprachwahrnehmung hat. Welche Befunde sprechen dafür?

Lektüreempfehlungen

- David B. Pisoni / Robert E. Remez (Hg.): *The Handbook of Speech Perception*, Malden 2005. *Umfassender und aktueller Überblick über die zentralen Bereiche der Sprachwahrnehmungsforschung.*
- Henning Reetz / Allard Jongman: *Phonetics: Transcription, Pro-duction, Acoustics and Perception*, Malden 2009. *Sehr gut ver-ständliche Einführung in die Teilgebiete der Phonetik mit Ab-schnitten zur Sprachwahrnehmung.*