

Christian Grüny  
Artikulation und Resonanz  
Sprachverstehen als zwischenleiblicher Vorgang

I.

»Zwei Menschen unterhalten sich in einer unbekanntem Sprache; ich höre zu. Heißt das, dass ich sie höre? Dieselben Schwingungen, die ihre Ohren treffen, kommen auch zu mir. Ich appetziere aber nur ein verworesenes Geräusch, wo ein Laut dem anderen ähnelt. Ich unterscheide nichts und könnte nichts wiederholen. Die beiden Redenden dagegen scheiden in dieser selben Schallmasse Konsonanten, Vokale und Silben, die recht verschieden sind, kurzum deutliche Worte. Wo ist der Unterschied zwischen ihnen und mir?«<sup>1</sup>

Diese Passage aus Bergsons *Materie und Gedächtnis* stellt immer noch eine der besten Formulierungen für das Problem dar, um das es mir hier gehen soll: Wie kann es sein, dass die einen da, wo die anderen nur ein diffuses akustisches Gemeinge wahrnehmen, artikulierte und verständliche Sprache hören? Das Problem liegt offen zutage, es bedarf nicht der wissenschaftlichen Reformulierung oder Verfröndung, um überhaupt als solches erkennbar zu sein. So prägnant aber die Situation selbst ist, die Bergson beschreibt, so sehr sie der Alltagserfahrung angehört, so wenig hat diese Erfahrung eine Lösung für das Problem anzubieten. Wer dann ist für seine Klärung zuständig? Die Sache, um die es hier geht, ist von einiger Tragweite: Wir haben es mit einem elementaren und dadurch besonders erhellenden Fall von Artikulation zu tun, der unmittelbar an leibliche Vollzüge gebunden und dennoch durch und durch kulturell geformt ist.

Das Verstehen sprachlicher Bedeutung ist seit vielen Jahrzehnten ein zentrales Thema philosophischer Auseinandersetzung gewesen, was man für das hier zu verhandelnde Problem nicht sagen kann. Die *speech recognition* im Gegensatz zur *language comprehension* ist viel mehr weitgehend in der Linguistik und den Kognitionswissenschaften behandelt worden, auch wenn sie alles andere als ein einzelwissenschaftliches Spezialthema ist. Das hat insofern seine Richtigkeit, als die Philosophie allein dieses Thema nicht bearbeiten kann, sondern auf die wissenschaftliche Forschung angewiesen bleibt. Wendet man sich dieser in unserem Fall zu, so trifft man auf eine einigmaßen kontroverse Diskussion und stellt dabei schnell fest, dass es ohne die Philosophie auch

<sup>1</sup> Henri Bergson, *Materie und Gedächtnis*, Hamburg: Meiner 1991, 101.

nicht gut geht – die entsprechenden Untersuchungen und theoretischen Modelle können eine philosophische und insbesondere eine phänomenologische Relektüre und Interpretation gut vertragen. Exemplarisch für eine solche Relektüre sind hier bis heute Bergsons und vor allem natürlich Merleau-Pontys Auseinandersetzung mit der Psychologie, Biologie, Neurologie, Linguistik etc. ihrer Zeit.

Ich möchte hier eine Theorie als Ausgangspunkt nehmen, die von Alvin Liberman und einigen Kollegen aus den Haskins Laboratories nach einigen Vorarbeiten 1967 zuerst prägnant formuliert worden ist: die sogenannte *motor theory of speech recognition*. Zuerst aber muss das Problem noch einmal schärfer gefasst werden, für das sie eine Lösung anbietet.

## 2.

Wir haben in der von Bergson beschriebenen Situation ein systematisches und ein damit zusammenhängendes empirisches Problem, wenn wir die Dinge denn einmal vorerst auf diese Weise sortieren wollen. Das systematische Problem hängt mit der für die Linguistik zentralen Unterscheidung zwischen Phonetik und Phonologie zusammen. Der Unterschied, nach dem Bergson fragt, lässt sich linguistisch so formulieren: Während er selbst nur Laute hört, hören die Sprecher Phoneme bzw. Phonemkombinationen. Ein Phonem ist bestimmt als kleinste bedeutungsdifferenzierende Einheit, also gerade nicht akustisch, aber auch noch nicht als Bedeutungsträger. Für den Hörer geht es um die Frage, an welcher Stelle er Identitäten und Differenzen anzusetzen hat – was überhaupt etwas ist, das zählt. Rein akustisch lässt sich dies nicht entscheiden, und daher liegt das Problem nicht darin, dass er die entscheidenden Stellen nicht kennt, an denen er im Lautstrom Schritte zu setzen hat. Um die lautliche Ebene zu gliedern, reicht eine entwickelte Sensibilität für reale Unterschiede eben nicht aus, sondern sie muss auf eine andere Ebene bezogen werden. Nur dann können da Identitäten aufgefasst werden, wo rein lautlich Unterschiede sind – man spricht hier von Allophonen –, und umgekehrt Differenzen erscheinen, wo auf der Lautebene das Gleiche vorliegt. Klassischerweise werden Phoneme daher als abstrakte Einheiten beschrieben, die im systematisch-differentiellen Gefüge der jeweiligen Sprache angesiedelt sind und nicht einzeln an realen lautlichen Unterschieden festgemacht werden können. Die Phonologie, die sich mit diesen Einheiten beschäftigt, kann als Ursprungsdisziplin des Strukturalismus gelten,<sup>2</sup> und sie ist der Bereich, in

dem an der strukturalistischen Grundfigur eines diakritischen Systems bis heute am plausibelsten festgehalten werden kann. Die Phoneme einer Sprache sind tatsächlich streng differenziell angeordnet und bestimmen sich gegenseitig; das phonologische System ist, ganz wie Saussure es formuliert hat, eine Form und keine Substanz.<sup>3</sup> Der produktive und weitgehend unproblematische Charakter der phonologischen Modelle lässt die Frage, ob es dieses System wirklich »gibt«, eher in den Hintergrund treten – anders etwa als bei Lévi-Strauss' elementaren Strukturen der Verwandtschaft oder gar Barthes' Sprache der Mode. Dafür stellt sich die philosophische Frage nach seinem Status, seiner Existenzweise, die den Phonologen bei seinen Untersuchungen nicht interessieren muss, hier auf besonders deutlich greifbare Weise: Wo ist dieses System?

Dazu kommt aber noch ein weiteres, diesmal die faktische Realisierung des Systems betreffendes Problem, das Liberman beschäftigt hat: Das Vorliegen von Allophonen ist nur die Spitze des Eisbergs, denn die Heterogenität des Lautlichen zum Phonologischen ist weit radikaler. Sieht man sich den realen Lautstrom an, so findet man derartig extreme Deformationen und Überblendungen, dass die Identifikation der gemeinten Phoneme endgültig als Rätsel erscheint. Weder gibt es zeitlich klar abgrenzbare Einheiten, noch lassen sich eindeutige Zuordnungsregeln zwischen dem Systematischen und dem Lautlichen angeben. Erklärungsbedürftig in der von Bergson beschriebenen Situation ist damit eher nicht, wieso er selbst nur undifferenzierte Klanglichkeit hört, sondern wieso die beiden Sprecher in dieser Klanglichkeit ihre Sprache vollkommen unmissverständlich auffassen können. Isoliert man einzelne Laute, etwa das /d/ in den Lautkombinationen /di/, /de/, /da/ und /du/, so erscheinen die Unterschiede radikal; die Sprecher einer Sprache hören davon aber nichts: »they hear the phonemes but not the intraphonemic variations.«<sup>4</sup> Man ist fast versucht zu sagen, dass das Lautliche zum größten Teil an ihnen vorbeigehet. Ihre Fähigkeit ist nicht nur eine der Differenzierung, sondern ebenso eine der gezielten Nivellierung.

Libermans Lösung für dieses Problem lautet wie folgt: Was in Wirklichkeit aufgefasst wird, ist gar nicht primär die lautliche Ebene, sondern das Artikulatorische, das sie erzeugt – »perception mirrors articulation more closely than sound.«<sup>5</sup> Die »Analyse« oder das »Decodieren« des Gehörten wäre damit nicht das Entziffern eines verschlüsselten Textes aufgrund genauer Kennnis der entsprechenden Codierungsregeln, son-

<sup>3</sup> Vgl. Ferdinand de Saussure, *Grundfragen der allgemeinen Sprachwissenschaft*, Berlin: de Gruyter 1967, 134.

<sup>4</sup> Alvin M. Liberman, Franklin S. Cooper, Donald P. Shankweiler u. Michael Studdert-Kennedy, »Perception of the speech code«, in: *Psychological Review* 74, 6 (1967), 431-461, hier 442.

<sup>5</sup> A. a. O., 453.

<sup>2</sup> Vgl. Nikolaj S. Trubetzkoy, *Grundzüge der Phonologie* (Travaux du Cercle Linguistique de Prague 7), Prag 1939.

dem der Nachvollzug einer Bewegung. Tatsächlich ist genau dies bereits Bergsons These gewesen. Die Basis dafür hatte er bereits in *Zeit und Freiheit* gelegt: Die zentrale, mit der Einführung der *durée* zusammenhängende These war dort, dass alle, auch quantitative Differenzen als Qualitäten und Qualitäten als Bewegungen aufgefasst werden müssen, so dass die Wahrnehmung insgesamt zu einem Gefüge differenzierter Vollzüge bzw. Nachvollzüge wird.<sup>6</sup> In *Materie und Gedächtnis* wird diese These schließlich auch auf die Sprache bezogen. Im Verstreuen einer Sprache wird Bergson zufolge keine Aufeinanderfolge von Lauten, sondern »das motorische Schema des gehörten Wortes«<sup>7</sup> aufgefasst, denn es gibt eine »Tendenz der auditiven Wortindrücke, sich in Artikulationsbewegungen fortzusetzen«<sup>8</sup> – sofern man die betreffende Sprache spricht. Eine Sprache zu verstehen, heißt damit immer, sie zu können.

Man könnte sich als phänomenologischer Leser kognitionswissenschaftlicher Forschung an dieser Stelle zufriedener geben: Mit einer Verspätung von mehreren Jahrzehnten und ohne Kenntnis der entsprechenden Vorarbeiten ist die Spracherkennungsforschung bei einer theoretischen Figur gelandet, die entscheidenden Einfluss auf die Phänomenologie der Leiblichkeit gehabt hat. Die motorischen Schemata, auf die Liberman und seine Kollegen gestoßen sind, sind leibliche Vorgänge, ja sie sind ein besonders unmittelbar anschauliches Beispiel zwischenleiblichen Geschehens – eine Indifferenzzone zwischen dem eigenen Tun und dem des Anderen, in dem beide nur wechselseitig durcheinander erfasst werden können. Unglücklicherweise aber ist diese Theorie weder in allen Punkten klar noch unkontrovers. Die Probleme liegen dabei sowohl in der Weise, wie sie von ihren Proponenten formuliert worden ist, als auch in der von verschiedener Seite vorgebrachten Kritik. Dass sie in jüngerer Zeit auf die Debatte um Spiegelneuronen bezogen worden ist, fügt einen weiteren schwierigen Punkt hinzu. Ich werde mich daher im Folgenden bemühen, auf einige dieser Punkte einzugehen und sie einer phänomenologischen Lesart zu unterziehen. Am Ende wird tatsächlich die These stehen, die hier bereits vorweggenommen wurde. Sie bedarf aber der klärenden Reformulierung.

## 3.

Die Verschiebung des Fokus von der erklarenden Sprache auf die motorischen Vorgänge provoziert bei aller Plausibilität doch eine Nachfrage: Impliziert die *Motor Theory*, dass es beim Sprachverstehen um die

6 Vgl. Henri Bergson, *Zeit und Freiheit*, Hamburg: EVA 1994, mit Anwendung auf Kunst und Musik 17 ff., 39 ff.

7 Bergson, *Materie und Gedächtnis*, a. a. O. (Anm. 1), 103, 8 A. a. O., 106.

Artikulation und *nicht* um das Lautliche geht? Und heißt das darüber hinaus, dass das gleiche für das Sprechen gilt? Liberman folgend müsste die Antwort zuerst einmal in beiden Fällen Ja sein: Entscheidend ist in beiden Fällen das Artikulatorische, wohingegen Varianten des Lautlichen, die dieses bis zur Unkennlichkeit verzerrten, vergleichsweise wenig ins Gewicht fallen. Kurz: Es kommt beim Sprechen wie beim Hören nicht auf das an, was man hört, sondern auf das, was man bzw. der andere tut. Für den sichtbaren Anteil der Artikulation ist sein Einfluss auf die Auffassung des Gesprochenen nachgewiesen worden; in Anlehnung an einen der beteiligten Forscher spricht man vom McGurk-Effekt, bei dem eine gesehene Artikulation die Auffassung einer gleichzeitig gehörten anderen Silbe grundlegend modifiziert, so dass die Probanden eine zwischen den beiden liegende Lautform wahrnehmen oder gar dem Gesehenen den Vorzug geben.<sup>9</sup> Wer einmal versucht hat, einem anderen etwas von den Lippen abzulesen, weiß allerdings, dass wir nur einen kleinen Teil dessen sehen, was eine Sprecherin tut, so dass wir das meiste nun doch über das Lautliche erschließen müssen. Wir hören sozusagen durch es hindurch, oder besser: Unser Hören aktiviert eine andere Bezugsebene, von der her es interpretiert wird.

Wenn man dies aber zugesteht, so stellt sich eine weitere Frage, die von den Kritikern immer wieder aufgeworfen wurde und die auch Liberman und seinen Kollegen natürlich bewusst ist: Wenn das Problem doch darin besteht, dass auf der lautlichen Ebene weder eine erkennbare Segmentierung noch eine klare Identifizierbarkeit der Einheiten anzutreffen ist, wieso sollte dies auf der artikulatorischen Ebene anders sein? Worauf sonst sollte der unterschiedliche Klang der Laute zurückzuführen sein, wenn nicht auf eine je verschiedene Artikulation?<sup>10</sup> Die mangelt dem Begriff der Koartikulation verhandelt, also der teilweise überlappenden Artikulation aufeinander folgender Phoneme – auch hier spricht wenig dafür, dass der Ebenenwechsel das Problem löst. Auf irgendeiner Ebene, so scheint es, muss sich eine tatsächliche Invarianz finden, die derjenigen der Phoneme entspricht; diese Ebene kann nicht die der realen Artikulationsbewegungen sein. Die *Motor Theory* löst dieses Problem mit einem weiteren Rückgang hinter die konkreten Vorgänge im Mund- und Rachenraum auf diejenige neuronaler Aktivität. Entsprechend wären danach die »motor commands«<sup>11</sup> als tatsächlich gegenüber

9 Vgl. Harry McGurk u. John MacDonald, »Hearing lips and seeing voices«, in: *Nature* 264 (1976), 746–748.

10 Vgl. etwa Dominic W. Massaro u. Trevor H. Chen, »The motor theory of speech perception revisited«, in: *Psychonomic Bulletin & Review* 15, 2 (2008), 453–457.

11 Liberman u. a., »Perception of the speech code«, a. a. O. (Anm. 4), 449.

den Verformungen der lautlichen und artikulatorischen Ebene neutrale, eindeutige Befehle. Sie wären der eigentliche Ort des Gemeinten, die konkrete Realisierung des Abstrakten.<sup>12</sup> Sie müssten als neuronale Repräsentation der Absichten der Sprecherin gelten, wie es in einer späteren Revision der Theorie auch ausdrücklich formuliert wurde: »[T]he objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands [...]«<sup>13</sup> Am Ende läuft dies auf eine Psychologisierung des phonologischen Systems hinaus, die nun auf die neuronale Ebene gestützt werden soll. Und es verschiebt die Frage nur, ohne sie wirklich zu beantworten: Wie haben die Hörer Zugang zu diesen motorisch-phonologischen Intentionen, wenn sie weder das sind, was man hört, noch das, was real im Mund der Sprecherin geschieht?

Die Antwort auf diese Frage hängt mit der zweiten zentralen These der *Motor Theory* zusammen, die sich auf die einfache Formel »speech is special« bringen läßt. Gemeint ist, dass das Sprachverstehen einen Mechanismus mobilisiert, der von allen anderen Wahrnehmungsvorgängen kategorial unterschieden und auf diese eine Funktion spezialisiert ist. Während zuerst noch relativ unspezifisch von einem »speech mode«<sup>14</sup> die Rede ist, wird dieser Wahrnehmungsmodus später mit einem Modul im Sinne Jerry Fodors identifiziert. Fodor bestimmt Module als bereicherspezifische, in sich abgeschlossene, dem Bewusstsein kaum bis gar nicht zugängliche Verarbeitungseinheiten des Geistes/des Gehirns.<sup>15</sup> Ein solches Modul soll nun auch für die Sprachverarbeitung zuständig sein und den Ort bereitstellen, an dem Lautlichkeit auf Artikulation und beide auf die als Intentionen bestimmten motorischen Schemata bezogen wird. Das Modul stelle einen Mechanismus dar, dessen Prinzip die einer Analyse durch Synthese ist, eines Erfassens durch virtuelles Nachahmen, es wäre ein »internal, innately specified vocal-tract synthesizer«<sup>16</sup>. Mit diesem Postulat muss die Frage nicht mehr lauten, wie wir Sprache verstehen können, sondern mit welchen Mechanismen das mentale System in uns ausgestattet sein muss, damit

12 Ganz in diesem Sinne heißt es: »Thus, motor commands are, in one sense, abstract >eme< type entities [...] in another sense [...] motor commands constitute the essential subset of real neural signals with which a general model of production and perception should be principally concerned.« (Ebd., Fn. 24)

13 Alvin M. Liberman u. Ignatius G. Mattingly, »The motor theory of speech perception revised«, in: *Cognition* 21 (1985), 1-36, hier 2.

14 Liberman u. a., Perception of the speech code, a. a. O. (Anm. 4), 443.

15 Vgl. Jerry A. Fodor, *The Modularity of Mind. An Essay in Faculty Psychology*, Cambridge u. London: MIT Press 1983, 47ff.

16 Liberman u. Mattingly, »The motor theory of speech perception revised«, a. a. O. (Anm. 13), 26.

es dies kann. Selbst die Integration anderer Wahrnehmungsmodi findet hier statt: Sie ist keine Funktion des Leiblich mit der Welt interagierenden Wesens, sondern des Moduls. Konsequenterweise findet sich bei einem Autor die Formulierung, dass dieses System »can lipread and perceive audio-visually«<sup>17</sup>.

Nun ist ein solches Modul neuronal nicht wirklich nachweisbar, und die Modularitätstheorie, die vor allem von Noam Chomsky mit seiner generativen Grammatik prominent vertreten wurde, hat insgesamt deutlich an Plausibilität verloren. Dafür ist in den vergangenen Jahren auf eine andere vielversprechende Entwicklung in den Neurowissenschaften zurückgegriffen worden: das Modell der Spiegelneuronen. Als die damit verbundenen Phänomene vor etwa zwei Jahrzehnten entdeckt wurden, wurde dies von manchen als neurowissenschaftliche Stützung der *Motor Theory* gewertet (auch wenn die Beeinflussung ursprünglich andersherum war – die Kognitionswissenschaftler, die zuerst mit der Spiegelneuronentheorie hervorgetreten sind, bezogen eine ihrer Inspirationen von Liberman): In den Gehirnen von Affen und, wie man mittlerweile weiß, auch von Menschen gibt es Neuronen, die nicht nur bei eigenen, sondern auch bei beobachteten Bewegungen – Greifbewegungen waren hier der erste Untersuchungsgegenstand – aktiv werden. Auf irgendeiner Ebene scheint zwischen eigenen und fremden Handlungen kein Unterschied gemacht zu werden – und war es nicht genau das, was die *Motor Theory* in Bezug auf das Sprachverstehen postuliert hatte? Bis heute herrscht allerdings nicht nur in der Theorie der Spiegelneuronen selbst, sondern gerade auch in ihrer Übertragung in unseren Kontext einige Unsicherheit, die von Kritikern schnell bemerkt wurde. Insbesondere die Modultese ist, wie sich schnell zeigen lässt, mit den Spiegelneuronen nicht kompatibel.<sup>18</sup>

Die gegenwärtige Diskussion ist davon abgekommen, diese Nervenzellen als »surprising new class of premotor neurons«<sup>19</sup> zu begreifen, also als besonderen Typ Neuronen, die im Gehirn klar verortet werden können und eine klar unrisse Fähigkeit bzw. Funktion haben. Vielmehr lassen sich die in den Experimenten aufgetretenen Zusam-

17 Quentin Summerfeld, »Visual Perception of Phonetic Gestures«, in: Ignatius G. Mattingly u. Michael Suddert-Kennedy (Hg.), *Modularity and the Motor Theory of Speech Perception*, Hillsdale, NJ: Erlbaum 1991, 117-137, hier 118.

18 Vgl. etwa Andrew J. Lotto, Gregory S. Hickok u. Lon L. Holt, »Reflections on mirror neurons and speech perception«, in: *Trends in Cognitive Science* 13 (2009), 110-114.

19 G. di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese u. G. Rizzolatti, »Understanding motor events: a neurophysiological study«, in: *Experimental Brain Research* 91 (1992), 176-180, hier 176.

menhänge zwischen eigenen und wahrgenommenen Handlungen in den unterschiedlichsten Gehirnregionen feststellen, so dass eher von einem Funktionsprinzip ausgegangen werden sollte als von einer Klasse von Neuronen. Keyzers und Gazzola haben dieses Prinzip als »shared circuits«<sup>20</sup> bezeichnet, also als prinzipiell gemeinsame Aktivierung der gleichen neuronalen Funktionszusammenhänge, die sich auf jede Form des Verhaltens und der Wahrnehmung erstreckt und auch Gefühle und Empfindungen umfasst; Wolfgang Prinz spricht in einem etwas anderen Kontext vom »common coding«-Prinzip.<sup>21</sup> Damit wird das Modell einer festen Verschaltung im Falle einer Gruppe von klar definierten Handlungen zugunsten einer weit allgemeineren Theorie verlassen, für die nun aber die Rede von Spiegelung nicht mehr wirklich angemessen erscheint: Die Spiegelmetapher impliziert eine automatische und strenge Replikation durch ein festes Medium und unterschätzt die Variabilität und Kontextrabhängigkeit – nur Handlungen, die zumindest im Prinzip selbst vollzogen werden können, die die Betreffende im Sinne des oben in Bezug auf die Sprache formulierten *kann*, aktivieren jene gemeinsamen Leitungsbahnen. Angemessener erscheint es mir daher, von Resonanz zu sprechen, weil damit sowohl der formale Zusammenhang als auch die konstruktive Differenz zwischen dem Eigenen und dem Beobachteten deutlich würde: Was resoniert, vollzieht eine in Bezug auf das ihm Begegnende zwar nicht beliebige, aber auch nicht durch dieses determinierte Bewegung.

In jedem Fall widerspricht diese Verallgemeinerung der »speech is special«-These diametral. Wenn die neuronale Grundlage des Sprachverstehens tatsächlich das Prinzip der kognitiv-motorischen Resonanz sein sollte, so wäre diese Funktion gerade nicht auf das Feld der Sprache beschränkt, und tatsächlich ist mittlerweile auch von Seiten der Vertreter der *Motor Theory* Zweifel an der Modultheorie angemeldet worden.<sup>22</sup> Damit würde sich die Theorie wiederum derjenigen Bergsons annähern, und erst dann wäre sie wirklich mit der Merleau-Pontyschen Vorstellung einer Zwischenleiblichkeit vereinbar.<sup>23</sup> Man muss sicher

20 Christian Keyzers u. Valeria Gazzola, »Towards a unifying neural theory of social cognition«, in: Silke Anders, Gabriele Ende, Markus Jungthofer, Johanna Kissler and Dirk Wildgruber (Hg.), *Progress in Brain Research* 156 (2006), 379-401.

21 Wolfgang Prinz, »Perception and Action Planning«, in: *European Journal of Cognitive Psychology* 9, 2 (1997), 129-154.

22 Vgl. mit explizitem Bezug auf die Spiegelneuronen- bzw. *shared circuits*-These Bruno Galantucci, Carol A. Fowler u. M. T. Turvey, »The motor theory of speech perception reviewed«, in: *Psychonomic Bulletin & Review* 13, 3 (2006), 361-377, insbes. 368 ff.

23 Vgl. dazu etwa Maurice Merleau-Ponty, »Der Philosoph und sein Schattenschein«, in: ders., *Zeichen*, Hamburg: Meiner 2007, 233-264.

aufpassen, dass die Verallgemeinerung der Spiegelneuronen zu *shared circuits* nicht zu vorzeitigen Übertragungen und Übergriffsgeräten führt; auch bedarf sie meines Erachtens einer phänomenologischen Kontextualisierung. Richtig verstanden kann sie aber tatsächlich für empirisch ansetzende Theoriearbeit mit philosophischer Relevanz genutzt werden, wie wir sie in unserem Beispiel vor uns haben. Ich kann sie an dieser Stelle nicht weiter diskutieren, sondern möchte mich auf der Grundlage der verallgemeinerten Resonanzfunktion, die sie postuliert, noch einmal den beiden oben angesprochenen Aspekten der *Motor Theory* zuwenden.

In beiden Fällen ging es um scheinbar kategoriale Trennungen: Artikulation vs. Lautlichkeit, invariante motorische Intentionen vs. verformte Realisierungen, und beide erscheinen zunehmend unplausibel. Die Position von Liberman und seinen Kollegen war zuerst einmal eindeutig: Es geht beim Sprachverstehen um Artikulation *und nicht* um Lautlichkeit. In einem späteren Text haben sie das Verhältnis der beiden Ebenen als »heteromorph« bezeichnet, denn es gehe um »objects whose dimensionalities are radically different from those of the proximal stimuli«<sup>24</sup>. Sie vergleichen dies mit der Wahrnehmung von Gegenständen im Raum vermittelt visuell Einzelinformationen wie glänzende Oberflächen etc. – also letztlich der Auffassung eines *Noemas* durch sensorische *Hyle* im Sinne des Husserl der *Ideen I*. Zu Recht wurde eingewandt, dass dies – wenn man es denn überhaupt so beschreiben will – für alle Wahrnehmung gilt.<sup>25</sup> Überdies ist ein empirisches Gegenargument schnell zur Hand: Wie erklärt man dann die erstaunlich erfolgreichen Anstrengungen, bei äußeren oder inneren Behinderungen des Sprechapparats die lautliche Ebene trotzdem so gut es geht konstant zu halten – und nicht möglichst die gleichen artikulatorischen Bewegungen zu vollziehen? Müsste es nicht andersherum sein?<sup>26</sup> Gerade diese Gegenüberstellung macht die Alternative als solche endgültig als irreführend erkennbar, und auch ihre Rekonstruktion im Sinne des transzendentalen Husserl macht die Sache nicht überzeugender.

24 Ignatius G. Mattingly u. Alvin M. Liberman, »Specialized Perceiving Systems for Speech and Other Biologically Significant Sounds«, in: Gerald M. Edelman, W. Einar Gall u. W. Maxwell Cowan (Hg.), *Auditory Function. Neurobiological Bases of Hearing*, New York u. a.: Wiley & Sons 1988, 775-793, hier 779 f.

25 Carol A. Fowler u. Lawrence D. Rosenblum, »The Perception of Phonic Gestures«, in: Ignatius G. Mattingly u. Michael Studdert-Kennedy (Hg.), *Modularity and the Motor Theory of Speech Perception*, Hillsdale, NJ: Lawrence Erlbaum 1991, 33-59.

26 Vgl. Lotto, Hickok u. Holt, »Reflections on mirror neurons and speech perception«, a. a. O. (Anm. 18), 113.

Entscheidend ist die Frage, wie wir Artikulation verstehen wollen. Es erscheint unproduktiv, sich damit auf die jeweils ganz konkreten Bewegungen der Stimmwerkzeuge zu beziehen, wenn man diese als messbare Vorgänge in der Welt befreit. Verloren geht dabei, dass sich mit dem Begriff keine beliebigen Bewegungsabläufe verbinden, sondern ein sinnhaft gegliedertes Verhalten. Von hier aus betrachtet erscheint es überaus eigenartig, bei der stimmlichen Artikulation eines lautlichen Zusammenhangs die beiden Dimensionen voneinander zu trennen und gegeneinander auszuspielen – in welche Richtung auch immer. Einen artikulierten Zusammenhang zu hören heißt nicht, durch eine reine Lautlichkeit hindurch auf eine Bewegung von Zunge und Kehlkopf gerichtet zu sein, sondern das Gehörte als artikuliertes aufzufassen. Insofern müsste die Theorie vielleicht besser als Artikulationstheorie denn als *Motor Theory* verstanden werden. Der Begriff, der mir dies am angemessensten zu beschreiben scheint, ist der der Geste.

## 4.

Gehen wir zuerst einmal von denjenigen Bewegungen aus, die wir im Alltagsverständnis am ehesten als Gesten bezeichnen: ausdruckshaften, hinweisenden oder auf sonstige Weise sinn durchdränkten Bewegungen von Armen und Händen. In Bezug auf eine ausholende, sanfte oder scharfe Geste erscheine eine kategoriale Trennung zwischen dem visuell wahrnehmbaren und den Betätigungen der verschiedenen Muskelgruppen, Knochen und Gelenke einigermaßen künstlich – sie sind leibliche Vollzüge. Das sinnhafte Moment bliebe bei einer solchen Trennung übrig und müsste einer weiteren noch zu identifizierenden Ebene zugeschlagen werden; naheliegenderweise wohl den dahinterstehenden Geist, mit dem der entziffernde Geist des Beobachters in ein Korrespondenzverhältnis treten müsste. Letztlich ist es genau diese den leiblichen Vollzug aufteilende Rekonstruktion, die die *Motor Theory* vorschlägt bzw. selbstverständlich übernimmt und durch die sie sich kaum lösbare Probleme einhandelt. Die wahrnehmende Auffassung von Gesten wird man ohne einen Nachvollzug nicht denken können, und hier zeigt sich noch einmal deutlich, dass wir es mit Resonanz- und nicht mit Spiegelungsverhältnissen zu tun haben: Zwar wird es nicht nur keine Schwierigkeit bereiten, sondern unvermeidbar sein, entsprechende Bewegungen überhaupt als gestisch aufzufassen. Der jeweilige Resonanzraum aber entscheidet darüber, ob bzw. besser zu welchem Grad ihr Sinn »verstanden« werden kann, vom widerstandslosen Nachvollzug unter Bekannten bis zur vollkommenen Dissonanz etwa angesichts mancher Tiere, denen wir dennoch gestisches Verhalten zuzuschreiben geneigt sind. Hier könnte, nebenbei bemerkt, auch ein kleiner Einwand gegen-

über Bergsons Beschreibung des Problems liegen: Auch derjenige, der nichts versteht, hört doch keinen diffusen Lautstrom, sondern gestisch artikulierende Bewegung, auch wenn er sie für ihn diffus und im Einzelnen nicht nachvollziehbar bleibt.

Auch Liberman und seine Kollegen hatten von »phonetic gestures« gesprochen, hatten diesem Begriff aber sozusagen nichts zugerannt, sondern ihn durch die zusätzliche Ebene der »motor commands« bzw. der neuronal verankerten Sprecherintentionen abgefangen. In der weiteren Diskussion sind theoretische Ansätze hervorgetreten, die an dieser Stelle angesetzt haben und ihn weiterentwickelt haben. Mit ihnen kann auch die Frage der Invarianz noch einmal angegangen werden, die den Rückgang auf die mentale bzw. neuronale Ebene ja motiviert hatte und die noch nicht wirklich beantwortet ist.

Man kann die ursprünglichen Einheiten der Sprachauffassung beim Kind als gestisch bezeichnen: Am Anfang steht der prosodische Verlauf, bei dem die expressive und bedeutungshafte Dimension nur sehr allmählich getrennt werden, und die erste Einheit ist das holistisch aufgefasste Wort.<sup>27</sup> Auch wenn diese ersten sprachlichen Größen gestisch sind, sind sie noch nicht wirklich artikuliert. Die ersten Worte werden sozusagen physiognomisch aufgefasst, mit durchaus unklarer innerer Gliederung, und dies zu einer Zeit, in der die eigene Sprechfähigkeit noch kaum entwickelt ist. Und sie fungieren holophrastisch, d. h. als Äquivalent einer ganzen Aussage oder als Marker einer bestimmten Situation bzw. eines Typs von Situation.

Das Lallen, das die Kinder in dieser Zeit produzieren, ist davon zuerst einmal vollkommen unabhängig; ja es lässt sich anfangs nicht einmal von der umgebenden Sprache beeinflussen. Wenn sie aber beginnen, Worte nachzusprechen, hat auf der Wahrnehmungsebene bereits eine Umstrukturierung von einer bloß akustischen Differenzierung nach Lauten zu einer protophonemischen Differenzierung nach Lauten der eigenen Sprache stattgefunden.<sup>28</sup> Am Anfang der eigenen Sprachproduktion aber steht das Wort als »articulatory program, or routine«<sup>29</sup>, also als komplexe, aber noch immer nicht untergliederte Geste, analog zur physiognomischen Auffassung. Studien zeigen, wie die Kinder sich

27 Vgl. Michael Studdert-Kennedy, »The phoneme as perceptuomotor structure«, in: A. Allport, D. Mackay, W. Prinz u. E. Scheerer (Hg.), *Language Perception and Production*, London: Academic Press 1987, 67-84, hier 76.

28 Vgl. Janet Werker, »The Ontogeny of Speech Perception«, in: Igenatius G. Martiny u. Michael Studdert-Kennedy (Hg.), *Modularity and the Motor Theory of Speech Perception*, Hillsdale, NJ: Lawrence Erlbaum 1991, 91-109.

29 A. a. O., 78.

von diesen wenigen Routineabläufen aus das Vokabular erschließen, indem sie sich zuerst an Worte halten, die diesen Abläufen folgen und sich allmählich weiter vorarbeiten – was hier soviel heißt wie in artikulatorisch immer entferntere Bereiche und dabei *gleichzeitig* immer differenzierter in die innere Gliederung der Worte. Die Aussprachefehler, die die Kinder in dieser Zeit machen, erscheinen aus dieser Perspektive wie systematische Verformungen, die von diesen Ausgangspunkten her rühren und von dort her typische Ersetzungen vornehmen, und ansonten vor allem das »phasing«, also die konkrete Abfolge der einzelnen Phoneme betreffen. Erst wenn sie die Gliederung nach phonemischen Gesten beherrschen, also *können*, steht ihnen potentiell das gesamte Vokabular der Muttersprache offen, das als Ansammlung konkreter lautlicher Physiognomien jedes Fassungsvermögen sprengen würde.

Erst jetzt ist das in Erscheinung getreten, was Martinet als »zweite Artikulation« bezeichnet hat<sup>30</sup>; die nochmalige Untergliederung der eigentlich bedeutungstragenden Momente in ein System von bedeutungsdifferenzierenden Einheiten. Was wir dann vor uns haben, ist eine Art Mikrogeste: winzige Bewegungsabläufe, die überdies für sich genommen keinerlei expressiven, appellativen oder hinweisenden Gehalt aufweisen, sondern lediglich differenzierend wirken. Inwiefern die Rede von Gesten dennoch auch hier gerechtfertigt erscheint, möchte ich abschließend zeigen.

Catherine Browman und Louis Goldstein haben ausgehend von diesen Mikrogesten eine »artikulatorische Phonologie« entwickelt; die das System phonologischer Kontraste endgültig in das tatsächliche Geschehen verlegt, das gestisch gedacht wird. Dabei ergeben sich Konsequenzen, die das Verhältnis von System und Wirklichkeit betreffen: »Since gestures have internal duration, they can overlap with each other; and since gestures are physical events, they are effected by physical processes occurring during the act of talking.«<sup>31</sup> Auch wenn sie es nicht ausdrücklich sagen, so scheint mir dabei doch entscheidend zu sein, dass es in diesem Modell keiner Invarianten mehr bedarf: Die abstrakte Konfrontation eines klar differenzierten Systems mit einer kontinuierlichen Wirklichkeit entfällt, weil letztere selbst gegliedert ist und diese Gliederung ausreicht, um für Identifizierbarkeit zu sorgen. Sie schließt Überlagerungen und spezifische Verformungen ein, die nun aber nicht auf einer kategorial anderen Ebene als die Einheiten selbst liegen. Das wäre gerade keine Segmentierung, sondern eine andere Form der Gliederung,

die eben gestisch-artikulatorisch ist. Ein Phonem als Geste wäre eine charakteristische Modulation, die sich sowohl auf den Stimmapparat als auch auf die hörbare Sprechweise beziehen lässt und die niemals in reiner Form vorliegt. Browman und Goldstein bringen zahlreiche Beispiele, wie auch Allophone als solche Verformungen begriffen werden können – sie sind lediglich so deutlich, dass sie den Phonologen auch schon früher aufgefallen sind.

Michael Studdert-Kennedy, der hier wichtige Arbeit geleistet hat, bezeichnet die neue Einheit gleichwohl wiederum als »a coordinative structure controlling an equivalence class of articulator movements, where equivalence is defined by function or goal, the achievement of a certain vocal tract configuration.«<sup>32</sup> Nimmt man aber das gestisch-leibliche Verständnis von Artikulation ernst, so muss man gerade keine höhere Koordinationsstruktur postulieren, die die Konfiguration der einzelnen Elemente des Sprechapparats zusammenhält. Funktionale Äquivalenz stellt sich unmittelbar her, indem die gestischen Vollzüge gar nicht erst von ihrem artikulatorischen Kontext losgelöst werden, sondern in ihren jeweiligen Formen als das je Gleiche aufgefasst werden. Insofern ist es auch irreführend, von Verformungen zu sprechen: Die je unterschiedliche Formung ist ihre Existenzweise. Ein komplexer Bewegungsablauf in der Interaktion mit der Welt und den Anderen bedarf, um gelernt und begriffen zu werden, keiner invarianten Repräsentation. Selbst die Rede von einem Schema ist schwierig, weil sie wiederum auf Invarianz setzt.

Die Feststellung, dass man Phoneme als solche nicht aussprechen kann, kann von hier noch einmal anders interpretiert werden: Unmöglich ist nicht dies deswegen, weil es sich um abstrakte Einheiten handelt, sondern weil es sie nicht in kontextfreier, reiner Form gibt. Ein Phonem artikulierbar heißt immer, es in einem sprachlichen Kontext anzusprechen. Bedeutungs-differenzierend sind diese gestischen Einheiten nur, wenn wir es tatsächlich mit Bedeutung, also mit Worten und Sätzen zu tun haben. Aber als Gefüge von Artikulationen des Gesichts sind sie real.

Resonanz ist der Modus, in dem diese Artikulationen gelernt und in dem sie verstanden werden. Als sprachliche Einheiten bilden sie ein weibliches Repertoire an Mikrogesten, das zuerst einmal niemandem zugerechnet werden kann. Wenn wir sie schließlich »haben«, so ist es sozusagen geborgtes Haben, das auf Bestätigung angewiesen ist und der Modifikation offensteht. Sie sind eine Form der Gliederung des Zwischenleiblichen.

30 Vgl. Andre Martinet, »La double articulation linguistique« in: *Travaux du Cercle Linguistique de Copenhague* 5 (1949), 30-37.

31 Catherine P. Browman u. Louis M. Goldstein, »Articulatory Phonology: An Overview«, in: *Haskins Laboratories Status Report on Speech Research* 1992, 23-42, hier 26 (dann in: *Phonetica* 49 (1992), 155-180).

32 Michael Studdert-Kennedy, »Mirror Neurons, vocal imitation, and the evolution of articulate speech«, in: Maxim I. Stamenov u. Vittorio Gallesse (Hg.), *Mirror Neurons and the Evolution of Brain and Language*, Amsterdam u. Philadelphia: John Benjamins 2002, 207-227, hier 212.

KULTUREN DER LEIBLICHKEIT

Herausgegeben von

Emmanuel Alloa, Thomas Bedorf,  
Christian Grüny und Tobias N. Klass

*»Kulturen der Leiblichkeit« ist die Publikationsreihe des gleichnamigen DFG-Netzwerks. Ihre Analysen verschiedener interdisziplinärer Themenfelder gelten der Leitfrage, was es für die Kulturwissenschaften bedeutet, zwischen Körperlichkeit und Leiblichkeit einen Unterschied zu machen.*

Band 1:

## Leib und Sprache

Zur Reflexivität verkörperter Ausdrucksformen

Herausgegeben von

Emmanuel Alloa und Miriam Fischer

# Leib und Sprache

Zur Reflexivität verkörperter Ausdrucksformen

Herausgegeben von

Emmanuel Alloa und Miriam Fischer

**VELBRÜCK  
WISSENSCHAFT**



# Inhalt

Publiziert mit Unterstützung des Schweizerischen Nationalfonds  
zur Förderung der wissenschaftlichen Forschung

Emmanuel Alloa und Miriam Fischer  
Leib und Sprache. Zur Einführung in ein verstricktes Thema . . . . . 7

## BEREDTE KÖRPER

Emil Angehrn  
Körper, Leib, Fleisch. Von den Inkarnationen der Sprache . . . . . 27

Joachim Küchenhoff  
Zwischenleiblichkeit und Körpersprache. Sinn und Nicht-Sinn  
körperbezogener psychischer Leiden . . . . . 45

Ludwig Jäger  
Die Leiblichkeit der Sprache. Phylogenetische Reminiszenzen  
in systematischer Absicht . . . . . 56

## KLANGKÖRPER UND ZEICHENTRÄGER

Christian Grünz  
Artikulation und Resonanz.  
Sprachverstehen als zwischenleiblicher Vorgang . . . . . 79

Andris Breiting  
Zeigende Rede. Welbezug und Gestaltungsspielraum  
der Sprache . . . . . 92

Mathias Flatscher  
Derrida – oder die Wiederkehr der Stimme . . . . . III

## VERLETZENDE WORTE

Steffen K. Herrmann  
Was heißt sprechen? Sozialität, Gewalt und Leiblichkeit  
der Sprache bei Pierre Bourdieu . . . . . 135

Gerald Posselt  
Die Gewalt der Tropen. Sprache, Schrift und Einschreibung  
bei Kafka und Nietzsche . . . . . 157

Erste Auflage 2013  
© Velbrück Wissenschaft, Weilerswist 2013  
www.velbrueck-wissenschaft.de  
Druck: Hubert & Co, Göttingen  
Printed in Germany  
ISBN 978-3-942393-60-7

### *Bibliografische Information der Deutschen Nationalbibliothek*

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der  
Deutschen Nationalbibliografie; detaillierte bibliografische Daten  
sind im Internet über <http://dnb.dtb.de> abrufbar.

Dieses Buch ist im Verlag Humanities Online  
([www.humanities-online.de](http://www.humanities-online.de)) als E-Book erhältlich.