

# Artikulatorische Sprachsynthese

Bernd J. Kröger

Institut für Phonetik der Universität zu Köln

## Zusammenfassung

In diesem Aufsatz wird Literatur zur Modellierung der Sprechtraktgeometrie, der Sprechtraktakustik und der artikulatorischen Steuerung zusammengestellt. Darüber hinaus werden kurz die Ergebnisse unserer Arbeiten zur Erweiterung des gestischen Steuerkonzeptes um ein prosodisches Modell skizziert.

## 1 Einleitung

Es existiert heute eine Vielzahl von artikulatorischen Sprachsynthesystemen. Allerdings werden unterschiedliche Ziele verfolgt. An den Bell Laboratorien (Coker 1968 und 1976, Flanagan et al. 1975 und 1980, Sondhi und Schroeter 1987, Schroeter und Sondhi 1992, Parthasarthy und Coker 1992) wird artikulatorische Synthese zur Sprachkodierung genutzt. An den Haskins Laboratorien (Mermelstein 1973, Rubin et al. 1981, Saltzman 1985, Saltzman und Munhall 1989, Browman und Goldstein 1990) werden Theorien zur Ansteuerung erforscht. In Stockholm (Fant 1960, 1972, 1982 und 1986, Fant et al. 1985, Liljencrants, 1985b, Lin 1990, Fant 1993) steht die Modellierung der Sprechtraktakustik und des glottalen Anregungssignals im Vordergrund. Schwerpunkt des in Leeds entwickelten Ansatzes (Allwood und Scully 1982, Scully 1990) ist die Modellierung der Aerodynamik. Beim Göttinger Modell (Strube 1977, Strube und Wilhelms 1982, Meyer et al. 1989, Fröhlich et al. 1994) wurde bereits sehr früh eine echtzeitfähige artikulatorisch basierte Sprachausgabe geschaffen. Eine sehr umfangreiche physiologisch orientierte Modellierung der Sprachproduktion wird am Massachusetts Institute of Technology geleistet (Perkell 1974, Wilhelms-Tricarico 1995).

Der Aufbau artikulatorischer Sprachsynthesysteme umfaßt die Modellierung der Sprechwerkzeuge und ihre Steuerung. Ein *Steuermodell* generiert die Artikulatorbewegungen. Die Modellierung der Sprechwerkzeuge umfaßt ein *Artikulatormodell*, das für jeden Zeitpunkt aus vorgegebenen Artikulatorpositionierungen die Formung des Sprechtraktes berechnet, und ein *akustisches Modell*, das aus vorgegebenen Sprechtraktformen das Sprachsignal generiert.

## 2 Die Sprechwerkzeuge

Das Hauptproblem bei der Entwicklung eines *Artikulatormodells* ist das Auffinden eines geeigneten Parametersatzes zur Beschreibung der möglichen Sprechtraktformen. Artikulatormodelle können auf einer direkten Parametrisierung der Ansatzrohrformung (Stevens und House 1955, Fant 1960, S. 63ff, Öhman 1967), auf einer Parametrisierung von Modellartikulatoren (Coker 1968 und 1976, Lindblom und Sundberg 1971, Mermelstein 1973, Lindblom et al. 1974, Heike 1979 und 1980), auf einer Faktorenanalyse mitsagittaler Konturen (Harshman et al. 1977, Maeda 1979, Gabioud 1994), auf physiologischen Analysen (Henke 1966, Perkell 1974) oder auf akustischen Analysen (Mrayati et al. 1988) basieren. All diesen Modellen gemeinsam ist, daß die Ansatzrohrformung nur mitsagittal, also nur zweidimensional beschrieben wird. Somit muß eine Transformation durchgeführt werden, die hieraus akustisch relevante Querschnittsflächenwerte des Ansatzrohres als Funktion des Ortes zwischen Glottis und Mund berechnet (Heinz und Stevens 1965, Sundberg et al. 1987, Perrier et al. 1992). Allerdings existieren auch erste Ansätze für dreidimensionale Artikulatormodelle (Fujimura 1977, Kakita et al. 1985, Wilhelms-Tricarico 1995).

*Akustische Modelle* können nach Frequenz- und Zeitbereichsverfahren unterteilt werden. Bei *Frequenzbereichsverfahren* wird die Übertragungsfunktion des Ansatzrohres aus dessen Geometrie berechnet und hiermit ein Formantsynthesator gesteuert (Fant 1960, Mathews und Walker 1962, Coker 1976, Badin und Fant 1984, Lin 1990, Parthasarthy und Coker 1992). Vorteil dieser Methoden ist die gute Modellierungsmöglichkeit frequenzabhängiger Verlustmechanismen im Sprechtrakt (z.B. Verluste durch Viskosität, durch Wärmeleitfähigkeit, durch Wandvibrationen, durch Abstrahlung an Lippen und Glottis). *Zeitbereichsverfahren* (Kelly und Lochbaum 1962, Maeda 1977 und 1982, Strube 1982, Liljencrants 1985, Meyer et al. 1989) erlauben die Generierung des Schallfeldes an jedem Ort im Ansatzrohr. Diese Verfahren erlauben auch die Modellierung der Sprechtrakt-aerodynamik. Die Vorteile beider Verfahren vereinigt ein hybrider Ansatz (Allen und Strong 1985, Sondhi und Schroeter 1987, Schroeter und Sondhi 1992).

Die primäre Anregung des Sprechtraktes kann über *selbstschwingende Glottismodelle* (Flanagan und Landgraf 1968, Ishizaka und Flanagan 1972, Titze 1973, 1974 und 1989, Titze und Talkin 1979, Childers et al. 1986 und 1987, Cranen und Boves 1987, Liljencrants 1991, Cranen und Schroeter 1995, Story und Titze 1995) oder über *parametrische Modelle* (Rosenberg 1971, Fant et al. 1985a) erfolgen. Vorteil der selbstschwingenden Glottismodelle ist die produktionsorientierte Definition der Steuerparameter in Form von Stimmlippenspannung und Stimmlippenabstand (glottale Ruheöffnungsfläche). Insbesondere werden hier auch segmental induzierte Änderungen des Grundfrequenzverlaufes (z.B. aufgrund glottaler Öffnungs- und Schließbewegungen oder aufgrund von Konstriktionsbildungen im Ansatzrohr) ohne die explizite Formulierung von Steuerregeln generiert. Allerdings sind die selbstschwingenden Glottismodelle noch nicht ausgereift genug, um die komplexen Vorgänge der Stimmlippen-schwingung insbesondere in Hinblick auf inter- oder intraindividuelle Stimmklangunterschiede zu modellieren.

Ebenso ist die Modellierung *sekundärer Schallquellen* (Fant 1960, S. 272ff, Flanagan et al. 1975, Shirai und Masaki 1983, Liljencrants 1985) - d.h. die Modellierung des im Sprechtrakt generierten Friktionsrauschens aufgrund turbulenter Luftströmung hinter Konstriktionen - in Hinblick auf die Amplituden/Zeit-Charakteristik und in Hinblick auf die spektrale Charakteristik noch nicht ausgereift. Hier besteht insbesondere ein Bedarf an experimentellen Daten (Shadle 1983 und 1985, Badin und Fant 1989). Die Realisierung primärer und sekundärer Schallquellen setzt darüber hinaus die Modellierung der Aerodynamik im Sprechtrakt und damit die Einbeziehung eines Modells des pulmonalen Bereichs (Ohala 1974 und 1990, Rothenberg 1968, Scully 1990) voraus.

## 3 Die Steuerung

Einerseits ist Kodierung des akustischen Signals in Form von *artikulatorischer Resynthese* möglich. Hier sind - wegen des Problems der Invertierung akustischer in artikulatorische Parameter (z.B. Strube 1974, Atal et al. 1978) - insbesondere auf neuronalen Netzwerken oder genetischen Algorithmen basierende Ansätze erfolgversprechend (Shirai und Kobayashi 1986, Schroeter und Sondhi 1992, Shirai 1993, Warneboldt und Strube

1993, McGowan 1994). Andererseits existieren parametrische *regelbasierte Steuermodelle*, die nach kinematischen und dynamischen Modellen unterschieden werden können. In *kinematischen Steuermodellen* werden für jede Lautrealisierung artikulatorische Targetwerte angesetzt und artikulatorische Transitionen zur Verbindung der Targets generiert (Heike 1980 und 1989, Scully 1987, Meyer et al. 1985, Meyer et al. 1989, Parthasarthy und Coker 1992). Bei *dynamischen Steuermodellen* hingegen liegt ein physikalisches Modell zur Generierung der Artikulatorbewegungen aufgrund von äußeren, auf die Artikulatoren wirkenden Kräften vor (Hiki und Oizumi 1974, Kakita und Hiki 1974, Coker 1976, Fujisaki 1977, Shigenaga und Ariizumi 1977, Saltzman und Munhall 1989, Bouabana und Maeda 1994, Perrier und Ostry 1994).

#### 4 Ein gestisch-prosodisches Modell der Steuerung

Sehr vielversprechend erscheint uns das gestische Konzept der Steuerung eines artikulatorischen Synthetisators (Saltzman und Kelso 1983, Saltzman 1985, Kelso et al. 1986, Saltzman und Munhall 1989). Aus der Sicht der Sprachproduktionsforschung können Gesten als zugrundeliegende Einheiten gesprochener Sprache angesehen werden (Fujimura 1981, 1986 und 1990). Dies wird insbesondere durch die Rückführung von diskreten segmentalphonetischen Reduktionsprozessen (z.B. Assimilation und Elision von Lauten bei schneller Sprechweise) auf kontinuierliche gestische Verschiebungsprozesse (Browman und Goldstein 1989, 1990 und 1992, Kröger 1993a) deutlich. Eine Geste kann als gerichtete Artikulatorbewegung mit dem Ziel der Ausbildung einer für eine Lautrealisierung wichtigen Konstriktion im Ansatzrohr (z.B. Ausbildung eines bilabialen Verschlusses zur Realisierung eines /b/, einer pharyngalen Enge zur Realisierung eines /a:/) definiert werden. Die Parameter zur quantitativen Beschreibung einer Geste sind die *Targetposition*, die das Ziel und damit die Richtung einer gestischen Artikulatorbewegung definiert, die *Eigenperiodendauer* (Federkonstante), die das Zeitintervall definiert, in dem die Targetposition (annähernd) erreicht wird, und die *Ablöse- und Assoziationsphase*, die den Realisierungsgrad und den Grad der zeitlichen Überlappung von Gesten definieren (Browman und Goldstein 1989 und 1990). Das gestische Modell wird zur Ansteuerung unseres artikulatorischen Synthetisators (Kröger 1993a und 1993b) benutzt. Ziel der zur Zeit durchgeführten Arbeiten ist die Erweiterung des gestischen Steuermodells um eine prosodische Komponente. Hierzu wurden zunächst artikulatorische Messungen zur Abschätzung dieser gestischen Parameter in Abhängigkeit von den prosodischen Faktoren Akzentuierung, Sprechtempo und Stellung durchgeführt (zum Meßverfahren siehe Kröger 1993c und 1994a, Oppen-Rhein et al. 1994 und Kröger et al. 1995). Die Ergebnisse können qualitativ wie folgt zusammengefaßt werden. Unterschiedliche Grade der *Akzentuierung* führen zur Änderung der Targetposition einer Geste. Mit zunehmendem Akzentuierungsgrad wird der Abstand zwischen der Neutralposition des aktivierten Artikulators und dem Target größer. Zunahme der Akzentuierung resultiert somit in einer Extremalisierung der Targetlagen, ohne aber die gestische Bewegungsrichtung zu ändern. Die Eigenperiodendauer kann sich in unterschiedlicher Weise ändern. Bedingung für die Änderung der Eigenperiodendauer zusammen mit der Änderung der Targetposition ist allerdings, daß mit Erhöhung des Akzentuierungsgrades keinesfalls eine Verringerung des artikulatorischen Aufwandes (Verringerung der gestischen Impulskosten nach Nelson 1983) einher geht. Die Änderung der *Sprechgeschwindigkeit* kann im gestischen Ansatz je nach Sprecher auf zwei unterschiedliche Arten modelliert werden. Strategie I: Bei Erhöhung der Sprechgeschwindigkeit ändern sich die Phasenrelationen derart, daß eine Zunahme der gestischen Überlappung und damit die Verringerung der zeitlichen Ausdehnung einer Geste erreicht wird. Alle anderen Parameter bleiben konstant. Strategie

II: Die Erhöhung der Sprechgeschwindigkeit wird durch Herabsetzung der Eigenperiodendauer bei konstanten Phasenrelationen und damit bei konstantem Grad gestischer Überlappung erreicht. Diese zweite Strategie kann allerdings auch eine Zunahme des artikulatorischen Aufwandes implizieren. *Finalität* wird durch die Erhöhung der Eigenperiodendauer erreicht. Zusätzlich kann hier auch eine Verringerung der zeitlichen Überlappung der Gesten hinzutreten.

Zur artikulatorischen Modellierung der Grundfrequenz- und Intensitätslevelkontur auf Satzebene nehmen wir an, daß die zugrundeliegenden artikulatorisch-prosodischen Steuerparameter der pulmonale Druck, der Stimmlippenabstand und die Stimmlippenspannung sind (Ohala 1990). Änderungen des subglottalen Drucks entstehen in erster Linie aufgrund von Änderungen des Stimmlippenabstandes wegen der aerodynamischen Wechselwirkung zwischen pulmonalem und glottalem System. Im Unterschied zu anderen Ansätzen wird der subglottale Druck hier somit nicht als Steuerparameter angenommen. Während die *Stimmlippenspannung* in erster Linie die *Grundfrequenz* determiniert, werden *Intensitätsleveländerungen* innerhalb einer Äußerung in erster Linie durch *Änderungen des Stimmlippenabstandes* bewirkt: Eine Erhöhung (Verringerung) des Stimmlippenabstandes bewirkt eine Verringerung (Erhöhung) des aerodynamischen Widerstandes der glottalen Konstriktion und bei konstantem pulmonalen Druck dann eine Verringerung (Erhöhung) des subglottalen Drucks und folglich eine Verringerung (Erhöhung) des akustischen Intensitätslevels. Das pulmonale System sorgt während einer Produktion einer Äußerung für annähernd konstanten pulmonalen Druck. Die Höhe des pulmonalen Drucks determiniert dabei die mittlere Lautstärke, mit der eine Äußerung produziert wird (Sprechstärke). Die konkrete Umsetzung dieser Theorie mittels eines artikulatorischen Sprachsynthesystems ist unter Einbeziehung eines Modells des pulmonalen Bereichs (Ohala 1974, 1976 und 1990) und unter Erweiterung des selbstschwingenden Zwei-Massen-Modells von Ishizaka und Flanagan (1972) um einen Bypass zur Modellierung des glottalen Abduktions- und Adduktionsverhaltens (Kröger 1994b) möglich. Die Ausmessung dieses Modells zeigt, daß bei einem (mittleren) pulmonalen Druck von 6cmH<sub>2</sub>O allein durch Variation von Stimmlippenspannung und Stimmlippenabstand ein Grundfrequenzbereich von rund einer Oktave und ein Dynamikbereich von über 10dB erzeugt werden kann. So kann dann beispielsweise der mit Akzentuierung verbundene Grundfrequenz- und Intensitätslevelanstieg mittels einer kombinierten glottalen Geste zur Anspannung der Stimmlippen und zur Verringerung des Stimmlippenabstandes realisiert werden.

#### 5 Diskussion und Ausblick

Aus der Sicht der Sprachsyntheseanwendungen hat artikulatorisch basierte Sprachsynthese heute kaum Bedeutung. Die Evaluierung unseres segmentalen artikulatorischen Sprachsynthesystems (Kröger und Oppen-Rhein 1992) zeigt, daß die Sprachqualität dieses Systems von der Qualität anderer heute gängiger Synthesysteme übertroffen wird (Fellbaum et al. 1994). Andererseits erreichen einige artikulatorisch basierte Synthetisatoren (Sondhi und Schroeter 1987, Lin 1990) aber bereits sehr hohe Sprachqualität zumindest im Fall artikulatorischer *Resynthese* vorgegebener akustischer Sprachbeispiele. Somit ist langfristig gesehen die Forschung im Bereich der artikulatorischen Sprachsynthese sinnvoll. Die bei artikulatorischer Sprachsynthese angestrebte detaillierte Nachbildung der menschlichen Sprachproduktion hat das Potential, die bei heute üblichen Sprachsyntheseverfahren methodisch bedingten Limitierungen der Sprachqualität (z.B. Modellierung von Koartikulation, Modellierung von Änderungen des Stimmklanges oder Modellierung von Segmentänderungen aufgrund von Sprechtempoerhöhung) zu überwinden.

## 6 Literatur

- Allen, D.R., Strong, W.J. (1985): "A model for synthesis of natural sounding vowels", *Journal of the Acoustical Society of America* 78, 58-69.
- Allwood, E., Scully, C. (1982): "A composite model of speech production", *Proceedings of the International Congress on Acoustics, Speech and Signal Processing ICASSP 82, Vol. 2*, 932-935.
- Atal, B.S., Chang, J.J., Mathews, M.V., Tukey, J.W. (1978): "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique", *Journal of the Acoustical Society of America* 63, 1535-1555.
- Badin, P., Fant, G. (1984): "Notes on vocal tract computation", *Speech Transmission Laboratory - Quarterly Progress and Status Report 2-3/1984* (Royal Institute of Technology, Stockholm), pp.53-107.
- Badin, P., Fant, G. (1989): "Fricative production modelling: Aerodynamic and acoustic data", *Proceedings of the 2nd European Conference on Speech Communication and Technology (EUROSPEECH '89), Vol. 2*, 23-26.
- Bouabana, S., Maeda, S. (1994): "Modélisation des mouvements articulatoires par la méthode de la LPC multi-impulsionnelle", *Journal de Physique 4, Colloque C5*, pp. 449-452.
- Browman, C.P., Goldstein, L. (1989): "Articulatory gestures as phonological units", *Phonology* 6, 201-251.
- Browman, C.P., Goldstein, L. (1990): "Tiers in articulatory phonology, with some implications for casual speech", in: J. Kingston, M.E. Beckman (Hrsg.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (Cambridge University Press, Cambridge), pp. 341-376. Auch in: *Haskins Laboratories Status Report on Speech Research SR-92* (1987), 1-30.
- Browman, C.P., Goldstein, L. (1992): "Articulatory phonology: An overview", *Phonetica* 49, 155-180.
- Childers, D.G., Alsaka, Y.A., Hicks, D.M., Moore, G.P. (1987): "Vocal fold vibrations: an EGG model", in: T. Bear, C. Sasaki, K.S. Harris (Hrsg.), *Laryngeal function in phonation and respiration* (Little Brown and Company, Boston), pp. 181-202.
- Childers, D.G., Hicks, D.M., Moore, G.P., Alsaka, Y.A. (1986): "A model for vocal fold vibratory motion, contact area, and the electroglottogram", *Journal of the Acoustical Society of America* 80, 1309-1320.
- Coker, C.H. (1968): "Speech synthesis with a parametric articulatory model", *Speech Symposium, Kyoto*, Paper A-4. Reprinted in: J.L. Flanagan, L.R. Rabiner (Hrsg.), *Speech synthesis* (Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1974), pp. 135-139.
- Coker, C.H. (1976): "A model of articulatory dynamics and control", *Proceedings of the IEEE* 64, 452-460.
- Cranen, B., Boves, L. (1987): "On subglottal formant analysis", *Journal of the Acoustical Society of America* 81, 734-746.
- Cranen, B., Schroeter, J. (1995): "Modeling a leaky glottis", *Journal of Phonetics* 23, 165-177.
- Fant, G. (1960): *Acoustic theory of speech production* (Mouton, Den Haag).
- Fant, G. (1972): "Vocal wall effects, losses, and resonance bandwidths", *STL-QPSR 2-3/1972*, pp. 28-52.
- Fant, G. (1982): "Preliminaries to analysis of the human voice source", *Speech Transmission Laboratory - Quarterly Progress and Status Report 4/1982* (Royal Institute of Technology, Stockholm), pp. 1-27.
- Fant, G. (1986): "Glottal flow: models and interaction", *Journal of Phonetics* 14, 393-399.
- Fant, G. (1993): "Some problems in voice source analysis", *Speech Communication* 13, 7-22.
- Fant, G., Liljencrants, J., Lin, Q. (1985a): "A four-parameter model of glottal flow", *Speech Transmission Laboratory - Quarterly Progress and Status Report 4/1985* (Royal Institute of Technology, Stockholm), pp. 1-13.
- Fant, G., Lin, Q., Gobl, C. (1985b): "Notes on glottal flow interaction", *Speech Transmission Laboratory - Quarterly Progress and Status Report 2-3/1985* (Royal Institute of Technology, Stockholm), pp. 21-45.
- Fellbaum, K., Klaus, H., Sotscheck, J. (1994): "Hörversuche zur Beurteilung der Sprachqualität von Sprachsynthesystemen für die deutsche Sprache", in: *Fortschritte der Akustik, Plenarvorträge und Fachbeiträge der 20. Jahrestagung für Akustik, DAGA'94* (Dresden), pp. 117-122.
- Flanagan, J.L., Landgraf, L.L. (1968): "Self-oscillating source for vocal-tract synthesizers", *IEEE Transactions in Audio and Electroacoustics AU-16*, 57-64.
- Flanagan, J.L., Ishizaka, K., Shipley, K.L. (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", *Bell System Technical Journal* 54, 485-506.
- Flanagan, J.L., Ishizaka, K., Shipley, K.L. (1980): "Signal models for low bit-rate coding of speech", *Journal of the Acoustical Society of America* 68, 780-791.
- Fröhlich, M., Warneboldt, H., Strube, H.W. (1994): "Berücksichtigung des subglottalen Systems bei der artikulatorischen Sprachsynthese", in: *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 19. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, DAGA '94* (Dresden), pp. 1305-1308.
- Fujimura, O. (1977): "Model studies of tongue gestures and the derivation of vocal tract area functions", in: M. Sawashima, F.C. Cooper (Hrsg.), *Dynamic aspects of speech production* (University of Tokyo Press, Tokyo), pp. 225-232.
- Fujimura, O. (1981): "Temporal organization of articulatory movements as a multidimensional phrasal structure", *Phonetica* 38, 66-83.
- Fujimura, O. (1986): "Relative invariance of articulatory movements: An iceberg model", in: J.S. Perkell, D.H. Klatt (Hrsg.), *Invariance and Variability in speech process* (Lawrence Erlbaum, Hillsdale, New Jersey), pp. 226-242.
- Fujimura, O. (1990): "Articulatory perspectives of speech production", in: W.J. Hardcastle, A. Marchal, *Speech production and speech modelling* (Kluwer Academic Press, Dordrecht), pp. 323-342.
- Fujisaki, H. (1977): "Functional models of articulatory and phonatory dynamics", in: R. Carre, R. Descout, M. Wajskop (Hrsg.), *Articulatory modelling and phonetics* (GALF, Group de la Communication Parlée, Grenoble), pp. 127-136.
- Gabioud, B. (1994): "Articulatory models in speech synthesis", in: E. Keller, J. Caelen (Hrsg.), *Fundamentals of speech synthesis and speech recognition. Basic concepts, state of the art and future challenges* (John Wiley, Chichester), pp. 215-230.
- Harshman, R., Ladefoged, P., Goldstein, L. (1977): "Factor analysis of tongue shapes", *Journal of the Acoustical Society of America* 62, 693-707.
- Heike, G. (1979): "Articulatory measurement and synthesis", *Phonetica* 36, 294-301.
- Heike, G. (1980): "Artikulatorische Synthese: Beschreibung des Programmpakets ARSDEM", *IPKöln-Berichte* 10, 8-31.
- Heike, G. (1989): "Ein phonologisches Artikulationsmodell des Deutschen", in: E. Feldbusch (Hrsg.), *Ergebnisse und Aufgaben der Germanistik am Ende des 20. Jahrhunderts: Festschrift für Ludwig Erich Schmitt zum 80. Geburtstag* (Olms-Weidmann, Hildesheim), pp. 159-173.
- Heinz, J.M., Stevens, K.N. (1965): "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech", *Proceedings of the Fifth International Congress of Acoustics, Paper A44*, (Liege, Belgium).
- Henke, W.L. (1966): *Dynamic articulatory model of speech production using computer simulation* (Dissertation, MIT,

- Cambridge, Massachusetts).
- Hiki, S., Oizumi, J. (1974): "Speech synthesis by rule from neurophysiological parameters", *Proceedings of the Speech Communication Seminar Stockholm 1974*, Vol. 2 (Almqvist & Wiksell, Uppsala), pp. 219-226.
- Ishizaka, K., Flanagan, J.L. (1972): "Synthesis of voiced sounds from a two-mass model of the vocal cords", *The Bell System Technical Journal* 51, 1233-1268.
- Kakita, Y., Hiki, S. (1974): "A study of laryngeal control for voice pitch based on anatomical model", *Proceedings of the Speech Communication Seminar Stockholm 1974*, Vol. 2, (Almqvist & Wiksell, Uppsala), pp. 45-54.
- Kakita, Y., Fujimura, O., Honda, K. (1985): "Computation of mapping from muscular contraction patterns to formant patterns in vowel space", in: V.A. Fromkin (Hrsg.), *Phonetic linguistics: Essays in honour of Peter Ladefoged* (Academic Press, Orlando), pp. 133-144.
- Kelly, J.L., Lochbaum, C.C. (1962): "Speech synthesis", *Proceedings of the International Congress on Acoustics*, Paper G-42, pp.1-4. Reprinted in: J.L. Flanagan, L.R. Rabiner (Hrsg.), *Speech synthesis* (Dowden, Hutchinson & Ross, Stroudsburg), pp. 127-130.
- Kelso, J.A.S., Saltzman, E.L., Tuller, B. (1986): "The dynamical perspective on speech production: Data and theory", *Journal of Phonetics* 14, 29-59.
- Kröger, B.J., Opgen-Rhein, C. (1992): "Ein phonembasiertes Produktionsmodell für artikulatorische Sprachsynthese", in: *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 18. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, DAGA '92* (Berlin), pp. 533-536.
- Kröger, B.J. (1993a): "A gestural production model and its application to reduction in German", *Phonetica* 50, 213-233.
- Kröger, B.J. (1993b): "A gestural approach for controlling an articulatory speech synthesizer", *Proceedings of the European Conference on Speech Communication and Technology, Eurospeech '93* (Berlin) 1903-1907.
- Kröger, B.J. (1993c): "Zur Analyse artikulatorischer Gesten mittels Articulographie-Daten", in: *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 19. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, DAGA '93* (Frankfurt), pp. 1044-1047.
- Kröger, B.J. (1994a): "Ein dynamischer Ansatz zur Parameterisierung von Sprechbewegungen", in: *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 20. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik DAGA '94* (Dresden), pp. 1309-1312.
- Kröger, B.J. (1994b): "Simulation of vocal fold oscillation behaviour by a self-oscillating glottis model", *Journal de Physique* 4 (Colloque 5, 3ème Congrès français d'acoustique, Toulouse), 457-460.
- Kröger, B.J., Schröder, G., Opgen-Rhein, C. (1995): "A gesture-based dynamic model describing articulatory movement data", *Journal of the Acoustical Society of America* 98, 1878-1889.
- Liljencrants, J. (1985): *Speech synthesis with a reflection-type line analog* (Dissertation, Royal Institute of Technology, KTH, Stockholm).
- Liljencrants, J. (1991): "A translating and rotating mass model of the vocal folds", *Speech Transmission Laboratory - Quarterly Progress and Status Report 1/1991* (Royal Institute of Technology, Stockholm), pp. 1-18.
- Lin, Q. (1990): *Speech production theory and articulatory speech synthesis* (Dissertation, Royal Institute of Technology, KTH, Stockholm).
- Lindblom, B., Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw, and larynx movements", *Journal of the Acoustical Society of America* 50, 1166-1179.
- Lindblom, B., Pauli, S., Sundberg, J. (1974): "Modeling coarticulation in apical stops", *Proceedings of the Speech Communication Seminar Stockholm 1974*, Vol. 2 (Almqvist & Wiksell, Uppsala), pp. 87-94.
- Maeda, S. (1977): "On a simulation method of dynamically varying vocal tract: Reconsideration of the Kelly-Lochbaum model", in: R. Carre, R. Descout, M. Wajskop (Hrsg.), *Articulatory modelling and phonetics* (GALF, Group de la Communication Parlée, Grenoble), pp. 281-287.
- Maeda, S. (1979): "An articulatory model of the tongue based on statistical analysis", *Journal of the Acoustical Society of America* 65, S22 (Suppl.1).
- Maeda, S. (1982): "A digital simulation method of the vocal-tract system", *Speech Communication* 1, 199-229.
- Mathews, M.V., Walker, P. (1962): "Program to compute vocal-tract poles and zeros", *Journal of the Acoustical Society of America* 34, p. 1977 (A)
- McGowan, R.S. (1994): "Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: Preliminary model tests", *Speech Communication* 14, 19-48.
- Mermelstein, P. (1973): "Articulatory model for the study of speech production", *Journal of the Acoustical Society of America* 53, 1070-1082.
- Meyer, P., Strube, H.W., Wilhelms, R. (1985): "Dynamische Anpassung eines artikulatorischen Modells an Lautübergänge", in: *Fortschritte der Akustik, Plenarvorträge und Fachbeiträge der 11. Tagung der Deutschen Arbeitsgemeinschaft für Akustik, DAGA '85*, pp. 611-614.
- Meyer, P., Wilhelms, R., Strube, H.W. (1989): "A quasiarticulatory speech synthesizer for German language running in real time", *Journal of the Acoustical Society of America* 86, 523-539.
- Mrayati, M., Carré, R., Guerin, B. (1988): "Distinctive regions and models: a new theory of speech production", *Speech Communication* 7, 257-286.
- Nelson, W.L. (1983): "Physical principles for economies of skilled movements", *Biological Cybernetics* 46, 135-147.
- Ohman, S.E.G. (1967): "Numerical model of coarticulation", *Journal of the Acoustical Society of America* 41, 310-320.
- Ohala, J.J. (1974): "A mathematical model of speech aerodynamics", in: *Reprints of the Speech Communication Seminar Stockholm 1974*, Vol. 2 (Almqvist & Wiksell, Uppsala), pp. 65-72.
- Ohala, J.J. (1976): "A model of speech aerodynamics", *Report of the Phonology Laboratory 1* (University of California, Berkeley), pp. 93-107.
- Ohala, J.J. (1990): "Respiratory activity in speech", in: W.J. Hardcastle, A. Marchal (Hrsg.), *Speech production and speech modelling* (Kluwer Academic Press, Dordrecht), pp. 23-54.
- Opgen-Rhein, C., Schröder, G., Kröger, B. J. (1994): "Zur Meßung dynamischer Parameter von Lippen-, Zungenspitzen- und Zungenrückengesten", in: *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 20. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, DAGA '94* (Dresden), pp. 1249-1252.
- Parthasarthy, S., Coker, C.H. (1992): "On automatic estimation of articulatory parameters in a text-to-speech system", *Computer Speech and Language* 6, 37-75.
- Perkell, J.S. (1974): *A physiologically-oriented model of tongue activity in speech production* (Dissertation, MIT, Cambridge, Massachusetts).
- Perrier, P., Boë, L.J., Sock, R. (1992): "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients", *Journal of speech and Hearing Research* 35, 53-67.
- Perrier, P., Ostry, D.J. (1994): "Dynamic modelling and control

- of speech articulators: application to vowel reduction", in: Keller, E., Caelen, J. (Hrsg.), *Fundamentals of speech synthesis and speech recognition. Basic concepts, state of the art and future challenges* (John Wiley, Chichester), pp. 231-252.
- Rosenberg, A. (1971): "Effect of glottal pulse shape on the quality of natural vowels", *Journal of the Acoustical Society of America* 49, 583-590.
- Rothenberg, M. (1968): *The breath-stream dynamics of simple-released-plosive production* (Karger, Basel).
- Rubin, P., Bear, T., Mermelstein, P. (1981): "An articulatory synthesizer for perceptual research", *Journal of the Acoustical Society of America* 77, 640-648.
- Saltzman, E.L. (1985): "Task dynamic coordination of the speech articulators: A preliminary model", *Haskins Laboratories Status Report on Speech Research SR-84*, 1-18.
- Saltzman, E.L., Kelso, J.A.S. (1983): "Skilled actions: A task dynamic approach", *Haskins Laboratories Status Report on Speech Research SR-76*, 3-50.
- Saltzman, E.L., Munhall, K.G. (1989): "A dynamic approach to gestural patterning in speech production", *Ecological Psychology* 1, 333-382.
- Schroeter, J., Sondhi, M.M. (1992): "Speech coding based on physiological models of speech production", in: S. Furui, M.M. Sondhi (Hrsg.), *Advances in speech signal processing* (Marcel Dekker, New York), pp.231-268.
- Scully, C. (1987): "Linguistic units and units of speech production", *Speech Communication* 6, 77-142.
- Scully, C. (1990): "Articulatory synthesis", in: W.J. Hardcastle, A. Marchal (Hrsg.), *Speech production and speech modelling* (Kluwer, Dordrecht), pp. 151-186.
- Shadle, C.H. (1983): "Turbulence noise in the vocal tract", *Proceedings of the 11th Congress in Acoustics, Vol. 4*, 171-174.
- Shadle, C.H. (1985): "The Acoustics of fricative consonants", *Technical Report 506* (Dissertation, MIT, Research Laboratory of Electronics, Cambridge, Massachusetts).
- Shigenaga, M., Ariizumi, H. (1977): "Articulatory movement by rule", in: R. Carre, R. Descout, M. Wajskop (Hrsg.), *Articulatory modelling and phonetics* (GALF, Group de la Communication Parlée, Grenoble), pp.193-202.
- Shirai, K. (1993): "Estimation and generation of articulatory motion using neural networks", *Speech communication* 13, 45-51.
- Shirai, K., Kobayashi, T. (1986): "Estimating articulatory motion from speech wave", *Speech Communication* 5, 159-170.
- Shirai, K., Masaki, S. (1983): "An estimation of the production process for fricative consonants", *Speech Communication* 2, 111-114.
- Sondhi, M.M., Schroeter, J. (1987): "A hybrid time-frequency domain articulatory speech synthesizer", *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-35*, 955-967.
- Stevens, K.N., House, A.S. (1955): "Development of a quantitative description of vowel articulation", *Journal of the Acoustical Society of America* 27, 484-493.
- Story, B.H., Titze, I.R. (1995): "Voice simulation with a body cover model of the vocal folds", *Journal of the Acoustical Society of America* 97, 1249-1260.
- Strube, H.W. (1974): *Bestimmung der Querschnittsfunktion des menschlichen Stimmkanals aus dem Sprachsignal* (Dissertation, Göttingen).
- Strube, H.W. (1977): "Synthesis part of a 'log area ratio' vocoder in analog hardware", *IEEE Transactions on Acoustics, Speech and Signal Processing ASSP-25*, 387-391.
- Strube, H.W. (1982): "Time-varying digital filter and vocal tract models", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 923-926.
- Strube, H.W., Wilhelms, R. (1982): "Synthesis of unrestricted German speech from interpolated log-area-ratio coded transitions", *Speech Communication* 1, 93-102.
- Sundberg, J., Johansson, C., Wilbrand, H., Ytterbergh, C. (1987): "From sagittal distance to area: a study of transverse, vocal tract cross-sectional area", *Phonetica* 44, 76-90.
- Titze, I.R. (1973): "The human vocal cords: a mathematical model. Part I", *Phonetica* 28, 129-170.
- Titze, I.R. (1974): "The human vocal cords: a mathematical model. Part II", *Phonetica* 29, 1-21.
- Titze, I.R. (1989): "A four-parameter model of the glottis and vocal fold contact area", *Speech Communication* 8, 191-201.
- Titze, I.R., Talkin, D.T. (1979): "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation", *Journal of the Acoustical Society of America* 66, 60-74.
- Warneboldt, H., Strube, H.W. (1993): "Bestimmung der akustisch-artikulatorischen Abbildung für ein Vokaltraktmodell", in: *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 19. Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, DAGA '93* (Frankfurt), pp. 1028-1031.
- Wilhelms-Tricarico, R. (1995): "Physiological modeling of speech production: Methods for modeling soft-tissue articulators", *Journal of the Acoustical Society of America* 97, 3085-3098.