ARTIKULATORISCHE SPRACHSYNTHESE

Bernd J. Kröger und Peter Birkholz

Klinik für Phoniatrie, Pädaudiologie und Kommunikationsstörungen, Universitätsklinikum Aachen und RWTH Aachen University
bkroeger@ukaachen.de, pbirkholz@ukaachen.de

Kurzfassung: Artikulatorische Sprachsynthese zielt darauf ab, den gesamten Prozess der Sprachproduktion ausgehend von der Generierung der Sprechbewegungen aus einer phonologisch-phonetischen Spezifikation einer zu generierenden Äußerung (Steuermodell) über die detaillierte Beschreibung der Positionierung, Formung und Funktion aller Artikulatoren (Lippen, Zunge, Unterkiefer, Gaumensegel, Kehlkopf, Lunge) zu jedem Zeitpunkt der Äußerung (Sprechtraktmodell) bis hin zur Generierung des akustischen Sprachsignals auf der Basis der geometrischen und funktionellen artikulatorischen Daten (akustisches Modell) zu modellieren. In diesem Artikel werden die zur Zeit Verwendung findenden Steuermodelle, Sprechtraktmodelle und akustischen Modelle beschrieben. Darüber hinaus werden potentielle Anwendungsgebiete für artikulatorische Sprachsynthese diskutiert.

Schlüsselwörter: Artikulatorische Sprachsynthese, Sprechtrakt, Artikulationsmodell, akustische Simulation, Phonation, Glottismodell, artikulatorische Steuerung, Sprechgeste, Artikulation, Phonetik.

1 Einleitung

Die Modellierung des gesamten Prozesses der Sprachproduktion ausgehend von der kognitiv-sensomotorischen Steuerung der Artikulationsorgane über die Generierung der Position und Oberflächenform der Artikulationsorgane und der Formung des Ansatzrohres für jeden Zeitpunkt einer Äußerung bis hin zur Generierung des akustischen Sprachsignals mittels Simulation der Schallausbreitung im Ansatzrohr kann unter dem Begriff artikulatorischer Sprachsynthese zusammengefasst werden. In den 70er Jahren des vorigen Jahrhunderts wurde Forschung zur artikulatorischen Sprachsynthese mit dem Ziel der Erreichung von Hochqualitätssynthese und damit der Reduzierung der Datenübertragungsraten bei Telefonie betrieben (Flanagan et al. 1980). Schnell erkannte man aber, dass die Idee der natürlichen Modellierung des menschlichen Sprachproduktionsprozesses nicht in einfacher Weise zu hochqualitativer Sprachsynthese führt. Zu viele Detailfragen zur artikulatorischen Steuerung wie auch zur Modellierung der Sprechtraktgeometrien und der akustischen Simulation waren ungeklärt. Ab den 80er Jahren wurde artikulatorische Sprachsynthese dann vor allem als phonetisches Forschungswerkzeug, z.B. für die Überprüfung artikulatorisch orientierter Hypothesen zur Sprachproduktion (siehe Artikulatorischen Phonologie/Phonetik: Saltzman and Munhall 1989, Browman and Goldstein 1992, Kröger 1993 und 1998) genutzt, da die zu dieser Zeit erreichbare Synthesequalität für die erfolgreiche Durchführung von Hörtests durchaus hinreichend war. In den letzten Jahren trat aber wieder das Ziel der Entwicklung hochqualitativer artikulatorischer Sprachsynthese in den Mittelpunkt des Interesses (z.B. Badin 2002, Birkholz et al. 2007). Diese Synthesetechnik könnte in den nächsten Jahren in all den Bereichen der Mensch-Maschine-Kommunikation nützlich sein, wo unterschiedliche Sprecher (unterschiedlicher „Stimmen“) gewünscht sind oder wo eine starke Variation des Grundfrequenzbereiches und der Stimmqualität (z.B. emotionales Sprechen, Sprechen und Singen) gewünscht wird. In diesem Beitrag soll der Stand der Forschung zur artikultorischen Sprachsynthese dargelegt werden und es sollen potentielle Anwendungsgebiete für artikulatorische Sprachsynthese vorgestellt werden.

2 Sprechtraktmodelle und die akustische Ebene

2.1 Sprechtraktmodelle

Sprechtraktmodelle generieren die Oberflächenform der beweglichen Artikulationsorgane (Zunge, Lippen, Unterkiefer, Gaumensegel, Stimmritze) und damit auch die Ansatzrohrformung als Funktion der Zeit auf der Basis einer definierten Menge von artikulatorischen Steuerparametern, d.h. auf der Basis von Parametern zur Beschreibung der Positionierung und Formung jedes Artikulators. Es können statistische, biomechanische und geometrische Sprechtraktmodelle unterschieden werden. *Statistische Sprechtraktmodelle* (z.B. Maeda 1988, Beautemps et al. 2001, Badin et al. 2002, Serrurier und Badin 2008) basieren auf artikulatorischen Datenkorpora gesprochener Sprache. Die geometrisch artikulatorischen Daten (mediosagittal zwei- oder auch dreidimensional) müssen zum Teil in manuellen Prozeduren „frame-by-frame“ erhoben werden. Die resultierenden artikulatorischen Steuerparameter können dann mittels statistischer Methoden (z.B. mittels Hauptkomponentenanalyse) abgeleitet werden. *Biomechanische Modelle* (z.B. Wilhelms-Tricarico 1995, Dang 2004) generieren die Formung und Bewegungen der Artikulationsorgane mittels eines zugrundeliegenden neuromuskulären Ansatzes. Physiologisches Wissen über Form, Beschaffenheit und die Funktion von Knochen, Knorpel, Muskeln und Gewebe wird hier in Finite-Elemente-Modellen zusammengefasst. Die artikulatorische Parametrisierung resultiert dann aus diesen physiologischen Vorgaben. *Geometrische Modelle* (z.B. Mermelstein 1973, Kröger 1998, Engwall 2003, Birkholz et al. 2006) nutzen aufgrund von phonetischem Wissen vorgegebene Parametrisierungen zur Formung und Positionierung der Artikulationsorgane (z.B. Absenkungswinkel des Unterkiefers, Höhe und Vor-/Rückverlagerung von Zungenrücken und Zungenspitze, Öffnungweite des Mundes, Rundungsgrad der Lippen, Grad der Absenkung des Gaumensegels, Öffnungsweite der Stimmritze). Geometrische Modelle sind flexibel und können an die physiologischen Rahmendaten beliebiger Sprecher unterschiedlichen Alters und unterschiedlichen Geschlechts angepasst werden (z.B. Birkholz und Kröger 2006).

2.2 Akustische Modelle

Die Aufgabe akustischer Modell ist es, auf der Basis der von den Sprechtraktmodellen generierten zeitabhängigen geometrischen Hohlrauminformationen die Schallausbreitung im Ansatzrohr und die Schallabstrahlung von Mund und Nase zu simulieren. Akustische Modelle können in Zeitbereichs- und Frequenzbereichsmodelle unterteilt werden. *Zeitbereichsmodelle* können in direkter Weise die an Impedanzsprüngen im Ansatzrohr entstehenden Reflexionen von vor- und rücklaufenden Teilwellen des Schalldruckes oder der Schallschnelle modellieren (r*eflektionsbasierte Modelle*: z.B. Kelly und Lochbaum 1962, Liljencrants 1985, Meyer et al. 1989, Kröger 1998) oder aber den Luftdruck und Luftstrom in definierten Teilstücken des Ansatzrohres aufgrund von schaltungsanalogen Netzwerken berechnen (*netzwerkbasierte Modelle*: z.B. Flanagan 1975, Maeda 1982, Birkholz et al. 2007). Der Vorteil der netzwerkbasierten Modelle ist, dass die Ortsdiskretisierung des Ansatzrohres nicht unbedingt in gleichlange Abschnitte erfolgen muss, und damit insbesondere die bei normalem Sprechen (z.B. in der Äußerung „au“) immer auftretende und akustisch nicht vernachlässigbare Längenvariation des Ansatzrohres so in einfacher Weise in die akustische Modellierung mit einbezogen werden kann. *Frequenzbereichsmodelle* (z.B. Allen und Strong 1985, Sondhi und Schroeter 1987) können insbesondere die aufgrund der akustischen Verlustmechanismen bei der Schallausbreitung im Sprechtrakt entstehenden Freqenzabhängigkeiten in der Übertragungsfunktion des Ansatzrohres in genauerer Weise nachbilden als dies in Zeitbereichsmodellen möglich ist. Allerdings ist der Rechenaufwand in Frequenzbereichsmodellen um ein Vielfaches höher als in Zeitbereichsmodellen, sodass heute zumeist netzwerkbasierten Zeitbereichsmodellen der Vorzug gegeben wird. Während die oben beschriebenen akustischen

Simulationen *leitungsanalog*, also räumlich eindimensional sind, gibt es mittlerweile auch Simulationsversuche zur Modellierung der dreidimensionalen Schallausbreitung im Sprechtrakt mittels Finite-Elemente-Methoden (z.B. El-Masri et al. 1996, Mazsuzaki und Motoki 2000).

2.3 Glottismodelle

Neben der Simulation der akustischen Schallausbreitung im Ansatzrohr muss auch die Schallentstehung an den schwingenden Stimmlippen (Primärschallentstehung) simuliert werden. Hier können selbstschwingende Glottismodelle und parametrische Glottismodelle unterschieden werden. *Selbstschwingende Glottismodelle* (z.B. Ishizaka und Flanagan 1972, Cranen und Bowes 1987, Story und Titze 1995, Kröger 1997a und 1997b, Alipour et al. 2000, Hunter et al. 2004) simulieren zunächst das mechanische Schwingungsverhalten der Stimmlippen aufgrund der aerodynamischen Anregung (sublottaler Druck und glottaler Luftstrom) und nachfolgend die Zeitfunktion des glottalen Luftstroms. Diese Modelle sind stark physiologisch orientiert (z.B. finite-Elemente-Modellierung der Struktur der Stimmlippen bei Alipur et al. 2000 und bei Hunter et al. 2004). Bei *parametrischen Glottismodellen* wird hingegen die Zeitfunktion der glottalen Öffnungsfläche (z.B. Titze 1989, Cranen und Schroeter 1996) oder direkt die Zeitfunktion des glottalen Volumenstroms (z.B. Fant et al. 1985, Fant 1993) vorgegeben. Der Vorteil der parametrischen Modelle ist, dass der Grundfrequenzverlauf und evtl. auch Teile des Stimmklanges in direkter Form determiniert werden können, während dies bei selbstschwingenden Glottismodellen nur indirekt über physiologische Parameter wie z.B. Ruheöffnungsfläche der Glottis bzw. aktiv voreingestellter (d.h. muskulär steuerbarer) Stimmlippenabstand und aktiv voreingestellte (d.h. muskulär steuerbare) Längsspannung der Stimmlippen erfolgt.

2.4 Modelle zur Rauschgenerierung

Neben der Schallentstehung an den Stimmlippen (Primärschallentstehung) wird insbesondere bei der Produktion von Plosiv- und Frikativlauten auch Rauschgenerierung im Ansatzrohr (Sekundärschallentstehung) genutzt. Auch hier können generisch-aerodynamische und parametrische Modelle unterschieden werden. *Generisch-aerodynamische Modelle* (z.B. Sinder 1999) generieren die Entstehung von Turbulenzen aufgrund der Geometrie des dreidimensionalen Sprechtraktes und er darin auftretenden Strömungsverhältnisse. Bei *parametrischen Modellen* (z.B. Mawass et al. 2000, Birkholz et al. 2007) hingegen wird ein Rauschen in das Ansatzrohr (zumeist bei eindimensionalen leitungsanalogen akustischen Modellen) eingespeist, dessen Intensität, dessen spektrale Färbung und dessen genauer Ort der Einspeisung auf der Basis der Geometrie der für die Rauschentstehung maßgeblichen Engebildung im Ansatzrohr und aufgrund der Stärke des Luftstroms berechnet wird. Auch hier werden zur Zeit weitgehend parametrischen Modelle gewählt, da eine dreidimensionale finite-Elemente-Simulation der gesamten Sprechtrakt-Akustik heute noch zu komplex und zu rechenaufwändig ist.

3 Modelle zur artikulatorischen Steuerung

Es ist die Aufgabe des Steuermodells, die zum Sprechen einer Äußerung nötigen Artikulationsbewegungen zu generieren. Dies umfasst nicht nur die Steuerung der supralaryngealen Artikulatoren (Zunge, Unterkiefer, Lippen, Gaumensegel) sondern auch die laryngeale Steuerung (aktiv voreingestellter Stimmlippenabstand und Stimmlippenspannung) und die sublaryngeale Steuerung (pulmonales Luftvolumen, aus dessen Änderung als Funktion der Zeit zusammen mit dem aerodynamischen Widerstand des laryngealen und supralaryngealen Systems der subglottale Druck resultiert). Für die Generierung von beliebigen Äußerungen

des Deutschen wurden zwei Steuermodelle, das segmentale Modell und das gestische Modell realisiert.

3.1 Segmentales Modell

Im *segmentalen Modell* (Kröger 1992, 1998 und 2003) wird für jeden Laut ein definiertes Zeitintervall (Produktionsintervall) angenommen. Innerhalb dieses Produktionsintervalls werden artikulatorische Zielmarken (Label) gesetzt. Zu den durch diese Marken definierten Zeitpunkten erreichen die artikulatorischen Steuerparameter definierte Werte und damit die zugehörigen Artikulatoren definierte Positionen. Beispielsweise wird für Vokale der Beginn und das Ende der Phonation und dazwischen das Erreichen der supralaryngealen vokalischen Zielform als Marke definiert; Im Fall eines stimmhaften Plosivlautes werden Marken für den Beginn und das Ende der supralaryngealen Verschlussbildung definiert. Koartikulation resultiert im segmentalen Modell aus artikulatorischer Unterspezifikation (Kröger 1998). So wird beispielsweise in der Silbe [bu:] und [bi:] für den Lippenschluss nur der Steuerparameter Mundöffnungsweite, nicht aber der Steuerparameter Grad der Lippenrundung spezifiziert, sodass schon während der konsonantischen Verschlussbildung im Fall des [bu:] die Lippenrundung und im Fall des [bi:] die Lippenspreizung eingeleitet bzw. fast vollständig ausgebildet werden kann.

3.2 Gestisches Modell

Im *gestischen Modell* (Kröger 1993 und 1998, Birkholz et al. 2006, Kröger und Birkholz 2007) werden artikulatorische Sprechgesten (speech gestures, vocal tract action units) als zugrundeliegende Einheiten der Artikulation angenommen (zur Theorie der Spechgeste siehe Saltzman und Munhall 1989, Browman and Goldstein 1992, Kröger 1993, Saltzman and Byrd 2000, Goldstein et al. 2006). Hier wird auf der phonologischen Ebene nach der Silbifizierung der Äußerung zunächst ein phonologisch-diskreter gestischer Plan spezifiziert, der dann unter Einbeziehung von extralinguistischen Faktoren wie Sprechsituation, Sprechtempo, Emotion etc. in einen phonetisch-quantitativen gestischen Plan und nachfolgend in einen artikulatorischen Plan überführt werden kann. Auf der Ebene des *phonologischen (oder diskreten) gestischen Plans* wird für jede Silbe eine Menge diskreter (oder phonologischer) Gesten spezifiziert. Beispielsweise wird für die Silbe /pa/ eine bilabiale Vollverschlussgeste, eine glottale Öffnungsgeste, eine dorsale Absenkgeste spezifiziert (Abb. 1a). In diesem Beispiel (wie in vielen anderen Fällen) können jedem Phonem eine oder mehrere Gesten zugeordnet werden. Es gibt aber auch Fälle wie z.B. /Spa/, wo für die beiden initalen stimmlosen Konsonanten nur eine gemeinsame glottale Öffnungsgeste spezifiziert wird. Darüber hinaus werden phonologische Sprechgesten miteinander assoziiert (Linien in Abb. 1a), d.h. es existiert ein Konzept, welche Geste auf im Zeitverlauf an die zeitliche Lage welcher übergeordneten Geste angebunden ist. So ist beispielsweise die glottale Öffnungsgeste in /pa/ an den Zeitpunkt der oralen Verschlusslösung gekoppelt (Abb. 1a). Auf der Ebene des *phonetischen (oder quantitativen) gestischen Plans* wird die zeitliche Ausdehnung jeder Sprechgeste und auch die für jede Geste zu spezifizierenden räumlich-artikulatorischen Zielpunkte quantitativ festgelegt (Abb. 1b). Während jede Geste normalerweise mehrere Artikulatoren ansteuert (z.B. wirkt eine labiale Verschlussgeste aktiv auf Ober-, Unterlippe und Unterkiefer ein), können nun aus dem phonetisch gestischen Plan (Abb. 1b) die detaillierten Bewegungsabläufe der einzelnen Modellartikulatoren mittels eines dynamischen Bewegungsmodells spezifiziert werden (artikulatorischer Plan, Abb. 1c). Hieraus lässt sich nachfolgend die Ansatzrohrform für jeden Zeitpunkt berechnen (Abb. 1d). Koartikulation entsteht in diesem Ansatz durch Koproduktion, d.h. durch die zeitliche und damit auch räumliche Überlappung von Sprechgesten. Darüber hinaus kann das gestische Modell in ein umfassendes kognitives und senso-

motorisches Gesamtmodell der Sprachproduktion, Sprachwahrnehmung und Sprachentwicklung eingebunden werden (Kröger et al. 2008 und 2009).

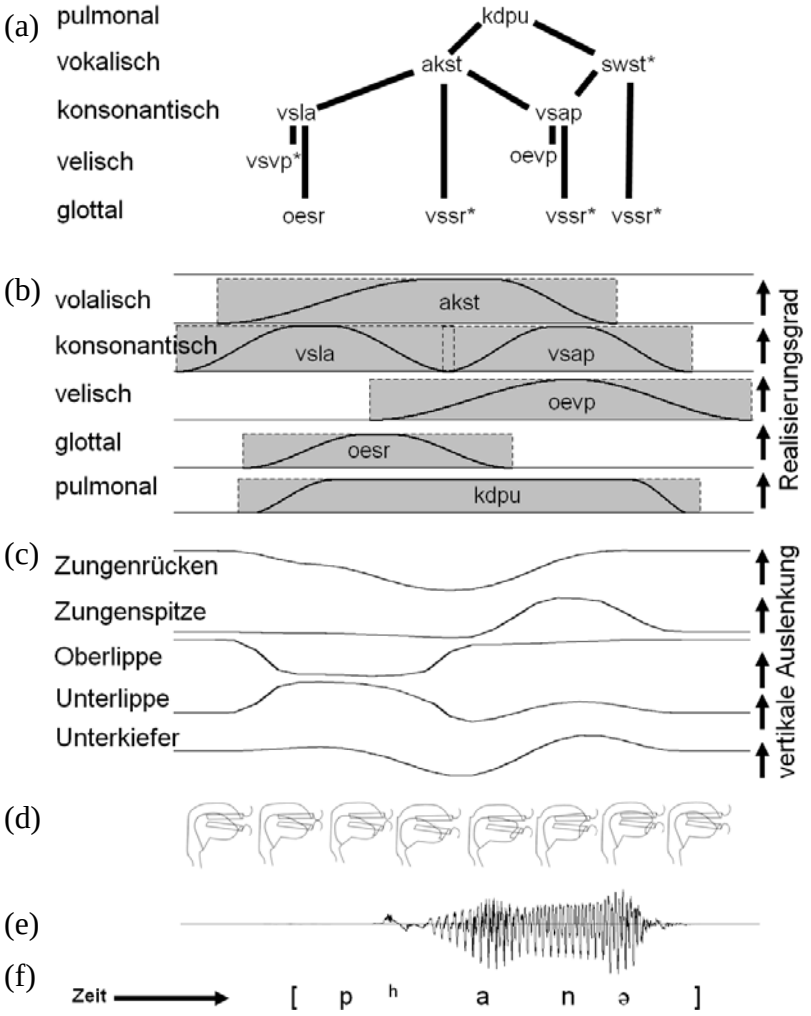

Abbildung 1: (a) Phonologisch-gestischer Plan, (b) phonetisch-gestischer Plan, (c) artikulatorischer Plan für ausgewählte Steuerparameter, (d) Mediosagittalschnitte zu ausgewählten Zeitpunkten, (e) akustisches Sprachsignal und (f) phonetische Transkription des Sprachsignals für das Wort „Panne“. (a) bis (e) wurde mittels des gestischen Modells von Kröger und Birkholz (2007) und des Sprechtrakt- und akustischen Modells von Birkholz et al. (2006) und (2007) generiert. Die gestischen Plane sind jeweils in 5 Schichten gegliedert: vokalisch, konsonantisch, velisch, glottal und pulmonal. Die Schicht „vokalisch“ enthält Gesten, die die gesamte suprapharyngeale Sprechtraktform spezifizieren. Die Schicht „konsonantisch“ enthält Gesten, die die konsonantische Artikulation einzelner Artikulatoren (Lippen, Zungenspitze oder Zungenrücken) spezifiziert. Die verbleibenden Schichten beinhalten jeweils die Sprechgesten des Gaumensegels, der Stimmritze und der Lunge. Für die Realisierung des Wortes „Panne“ werden benötigt: Geste zur Realisierung eines pulmonal konstanter Drucks (kdpu), Geste zur Realisierung der Sprechtraktform eines kurzen /a/ bzw. eines Schwalautes (akst, swst), Geste zur Realisierung einer labialen bzw. apikalen Verschlussbildung (vsla, vsap), Geste zur Realisierung einer Verschließung bzw. Öffnung der velopharyngealen Pforte (vsvp, oevp), Geste zur Realisierung einer Öffnung bzw. Verschließung der Stimmritze (oesr, vssr). Mit Stern gekennzeichnet sind die drei default-spezifizierte Gesten. Diese Gesten realisieren einen nichtnasalen Schwa-Laut bei normaler Phonation und treten im phonetisch-gestischen Plan nicht auf, da sie den Neutralzustand des Systems ohne gestische Aktivität definieren. Diese Gesten kontrollieren das System immer dann, wenn keine anderen Gesten auf der entsprechenden Schicht (vokalisch, velisch oder glottal) aktiv sind (siehe auch Kröger 1998).

4 Anwendungsgebiete

Wie in der Einleitung beschrieben, kann artikulatorische Sprachsynthese heute noch nicht als Hochqualitätssynthese ähnlich wie korpusbasierte Sprachsynthese (Iida et al. 2003) eingesetzt werden. Dazu ist die Qualität der zur Zeit verfügbaren Systeme noch nicht ausreichend. Als wichtigster Punkt ist ein noch existierendes Defizit in der Nachbildung der artikulatorischen Bewegungsdynamik zu nennen. Es ist derzeit noch nicht möglich, die dreidimensionale Formung des Ansatzrohres in ausreichender zeitlicher und räumlicher Auflösung in vivo zu messen. Darüber hinaus ist auch die Sekundärschallgenerierung als kritischer Punkt im Bereich des akustischen Modells bei artikulatorischer Sprachsynthese zu nennen. Bereits typische Anwendungen sind aber neben der Nutzung von artikulatorischer Sprachsynthese als phonetisches Forschungswerkzeug (z.B. Kröger 1993 und 1998 zur Beschreibung von segmentalen Reduktionsphänomenen im Deutschen) der Einsatz von zwei- und dreidimensionalen Artikulationsmodellen in interaktiver Software zum Aussprachetraining (Badin et al. 2008, Engwall et al. 2006). Zwar ist zwischen der akustischen Qualität der Systeme aus den 90er Jahren des letzten Jahrhunderts (z.B. Kröger 1998) und den heute realisierten Systemen (z.B. Birkholz et al. 2006) ein großer Qualitätssprung hörbar, aber erst eine noch weitere Qualitätsverbesserung hinsichtlich der artikulatorischen Steuerung und der Rauschgenerierung wird es sinnvoll erscheinen lassen, artikulatorisch basierte Systeme auch in typischen Spachsynthese-Anwendungsgebieten zu etablieren. Der Vorteil artikulatorischer Sprachsynthesesysteme wird dann insbesondere in ihrer Flexibilität liegen. Insbesondere ist dann die Modellierung unterschiedlicher Sprecher und unterschiedlicher Sprechstile einschließlich unterschiedlicher emotionaler Zustände und eine starke Variation des Grundfrequenzbereiches und des Sprechtempos ohne Qualitätseinbuße möglich.

Danksagung

Diese Arbeit wurden zum Teil aus Mitteln der Deutschen Forschungsgemeinschaft (DFG Projekt Nr. Kr 1439/15-1) gefördert.

Literatur

Alipour, F., Berry, D.A., Titze, I.R.: A finite-element model of vocal-fold vibration. Journal of the Acoustical Society of America 108 (2000) 3003-3012

Allen, D.R., Strong, W.J.: A model for synthesis of natural sounding vowels. Journal of the Acoustical Society of America 78 (1985) 58-69

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C., Savariaux, C.: Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. Journal of Phonetics 30 (2002) 533-553

Badin P, Tarabalka Y, Elisei F, Bailly G: Can you “read tongue movements”? Proceedings of Interspeech 2008 (Brisbane, Queensland, Australia, 2008) pp. 2635-2638

Beautemps, D., Badin, P., Bailly, G.: Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. Journal of the Acoustical Society of America 109 (2001) 2165-2180

Birkholz, P., Kröger, B.J.: Vocal tract model adaptation using magnetic resonance imaging. Proceedings of the 7th International Seminar on Speech Production. Belo Horizonte, Brazil (2006) 493-500

Birkholz, P., Jackèl, D., Kröger, B.J.: Construction and control of a three-dimensional vocal tract model. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006) Toulouse, France (2006) 873-876

Birkholz, P., Jackèl, D., Kröger, B.J.: Simulation of losses due to turbulence in the time-varying vocal system. IEEE Transactions on Audio, Speech, and Language Processing 15 (2007) 1218-1225

Browman, C.P., Goldstein, L.: Articulatory phonology: An overview. Phonetica 49 (1992) 155-180

Cranen, B., Boves, L.: On subglottal formant analysis. Journal of the Acoustical Society of America 81 (1987) 734-746

Cranen, B., Schroeter, J.: Physiologically motivated modeling of the voice source in articulatory analysis / synthesis. Speech Communication 19 (1996) 1-19

Dang, J., Honda, K.: Construction and control of a physiological articulatory model. Journal of the Acoustical Society of America 115 (2004) 853-870

El-Masri, S., Pelorson, X., Saguet, P., Badin, P.: Vocal tract acoustics using the transmission line matrix (TLM) method. Procedings of ICSPL, Philadelphia (1996) 953-956

Engwall, O.: Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. Speech Communication 41 (2003) 303-329

Engwall O, Bälter O, Öster AM, Kjellström H: Designing the user interface of the computer-based speech training system ARTUR based on early user tests. Journal of Behaviour and Information Technology 25 (2006) 353-365

Fant, G.: Some problems in voice source analysis. Speech Communication 13 (1993) 7-22

Fant, G., Liljencrants, J., Lin, Q.: A four-parameter model of glottal flow. Speech Transmission Laboratory - Quarterly Progress and Status Report 4/1985. Royal Institute of Technology, Stockholm (1985) 1-13

Flanagan, J.L., Ishizaka, K., Shipley, K.L.: Synthesis of speech from a dynamic model of the vocal cords and vocal tract. The Bell System Technical Journal 54 (1975) 485-506

Flanagan, J.L., Ishizaka, K., Shipley, K.L.: "Signal models for low bit-rate coding of speech. Journal of the Acoustical Society of America 68 (1980) 780-791

Goldstein, L., Byrd, D., Saltzman, E.: The role of vocal tract action units in understanding the evolution of phonology. In: M.A. Arbib (ed.) Action to Language via the Mirror Neuron System. Cambridge University Press, Cambridge (2006) 215-249

Hunter, E.J., Titze, I.R., Alipour, F.: A three-dimensional model of vocal fold abduction/adduction“ Journal of the Acoustical Society of America 115 (2004) 1747-1757

Iida, A., Campbell, N., Higuchi, F., Yasumura. M.: A corpus-based speech synthesis system with emotion. Speech Communication 40 (2003) 161–187

Ishizaka, K., Flanagan, J.L.: Synthesis of voiced sounds from a two-mass model of the vocal cords. The Bell System Technical Journal 51 (1972) 1233-1268

Kelly, J.L., Lochbaum, C.C: Speech synthesis. In: J.L. Flanagan, L.R. Rabiner (eds.) Speech Synthesis. Dowden, Hutchinson & Ross, Stoudsburg (1962) 127-130

Kröger, B.J.: Minimal rules for articulatory speech synthesis. In: J. Vandewalle, R. Boite, M. Moonen, A. Oosterlinck (eds.) Signal Processing VI: Theories and Applications. Elesevier, Amsterdam (1992) 331-334

Kröger, B.J.: A gestural production model and its application to reduction in German. Phonetica 50 (1993) 213-233

Kröger, B.J.: Zur artikulatorischen Realisierung von Phonationstypen mittels eines selbstschwingenden Glottismodells. Sprache-Stimme-Gehör 21 (1997a) 102-105

Kröger, B.J.: On the quantitative relationship between subglottal pressure, vocal cord tension, and glottal adduction in singing. Proceedings of the Institute of Acoustics 19 (1997b) 479-484 (ISMA97)

Kröger, B.J.: Ein phonetisches Modell der Sprachproduktion. Niemeyer Verlag, Tübingen (1998) siehe auch: www.speechtrainer.eu

Kröger, B.J.: Ein visuelles Modell der Artikulation. Laryngo-Rhino-Otologie 82 (2003) 402-407

Kröger, B.J., Birkholz, P.: A gesture-based concept for speech movement control in articulatory speech synthesis. In: Esposito, A., Faundez-Zanuy, M., Keller, E., Marinaro, M. (eds.): Verbal and Nonverbal Communication Behaviours. LNAI 4775, Springer Verlag, Berlin Heidelberg New York (2007) 174-189

Kröger, B.J., Lowit, A., Schnitker, R.: The organization of a neurocomputational control model for articulatory speech synthesis. In: A. Esposito, N. Bourbakis, N. Avouris, I. Hatzilygeroudis (eds.) Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. LNAI 5042, Springer Verlag, Berlin (2008) 126-141

Kröger, B.J., Kannampuzha, J., Neuschaefer-Rube, C.: Towards a neurocomputational model of speech production and perception. Speech Communication 51 (2009) 793-809

Liljencrants, J.: Speech Synthesis with a Reflection-Type Line Analog. Dissertation, Royal Institute of Technology, Stockholm (1985)

Maeda, S.: A digital simulation of the vocal-tract system. Speech Communication 1 (1982) 199-229

Maeda, S.: An articulatory model based on statistical analysis. Journal of the Acoustical Society of America 84, Supl.1 (1988) p. 146

Matsuzaki, H., Motoki, K.: FEM analysis of 3D vocal tract model with asymmetrical shape. Proceedings of the 5th Seminar on Speech Production. Seeon, Germany (2000) 329-332

Mawass, K., Badin, P., Bailly, G.: Synthesis of French Fricatives by Audio-Video to Articulatory Inversion. Acta Acustica, 86 (2000) 136-146

Mermelstein, P.: Articulatory model for the study of speech production. Journal of the Acoustical Society of America 53 (1973) 1070-1082

Meyer, P., Wilhelms, R., Strube, H.W.: A quasiarticulatory speech synthesizer for German language running in real time. Journal of the Acoustical Society of America (1989) 86: 523-540

Saltzman, E.L., Munhall, K.G.: A dynamic approach to gestural patterning in speech production. Ecological Psychology 1 (1989) 333-382

Saltzman, E., Byrd, D.: Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. Human Movement Science 19 (2000) 499-526

Serrurier, A., Badin, P.: A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. Journal of the Acoustical Society of America 123 (2008) 2335-2355

Sinder, D.J.: Speech synthesis using an aeroacoustic fricative model. PhD thesis, Rutgers University, New Jersey (1999)

Sondhi, M.M., Schroeter, J.: A hybrid time-frequency domain articulatory speech synthesizer. IEEE Transactions on Acoustics, Speech, and Signal Processing 35 (1987) 955-967

Story, B.H., Titze, I.R.: Voice simulation with a body cover model of the vocal folds. Journal of the Acoustical Society of America 97 (1995) 1249-1260

Titze, I.R.: A four-parameter model of the glottis and vocal fold contact area. Speech Communication 8 (1989) 191-201

Wilhelms-Tricarico, R.: Physiological modelling of speech production: Methods for modelling soft-tissue articulators, Journal of the Acoustical Society of America 97 (1995) 3085-3098