ARTICULATORY SPEECH SYNTHESIS IN THE CONTEXT OF SPEECH RESEARCH AND SPEECH TECHNOLOGY: REVIEW AND PROSPECT

Bernd J. Kröger[1]

[1]*Department for Phoniatrics, Pedaudiology, and Communication Disorders, Medical School, RWTH Aachen University, Aachen, Germany*
bernd.kroeger@rwth-aachen.de

Abstract: Articulatory speech synthesis is currently used mainly in speech and language research, while technical applications of speech synthesis are realized using corpus-based synthesis methods. From the point of view of speech and language research, the development of neurobiologically oriented approaches for modeling the entire speech processing system, i.e., modeling of speech perception (comprehension) as well as of speech production is challenging and of great interest. Here, cognitive-linguistic as well as sensorimotor aspects of production and perception need to be modelled and a biomechanical articulation model including neuromuscular control should be included as front-end module. From the viewpoint of speech and language technology, high-quality articulatory synthesis is realized mainly by corpus-based synthesis methods but also the use of geometric-parametric articulation models in conjunction with aeroacoustic vocal tract models seems to deliver promising results. This review article tries to summarize the major steps in research and in technical development of articulator speech synthesis. Furthermore, prospective, and reachable goals will be discussed for articulatory speech synthesis as a basic research tool and as technological speech synthesis approach.

1 Introduction

Copying the human speech apparatus and the mimicry of human speech even by using specific simplifications has challenged scientists for centuries [1]. Early mechanical as well as early computer-implemented simulation models of the vocal folds [2], of the vocal tract (for first geometrically based articulatory models see [3, 4]) and early acoustic models of the vocal tract (e.g., [5]) served as excellent tools for answering fundamental questions concerning sound generation in the vocal tract, concerning the formation of vocal tract resonances (formants) and concerning the alteration of formant frequencies over time as a result of articulation movements. But a lot of detailed knowledge was and is still missing.

During the decades of analogue and later of digital telephony technicians all over the world were confronted with the fact of a rather limited transmission infrastructure. This fact constituted a pressure on research groups and on industrial developer groups to find effective speech signal coding strategies. The resulting research efforts took place in parallel and in some labs also in synchrony, i.e., as part of research in digital articulatory speech synthesis during the late 60s and then further during the beginning 70s of the last century [6-8]. However, even before the turn of the millennium it was stated that solving the problem of developing high quality and natural articulatory speech synthesis still represents a challenge (see [9], p. 222): "It has been hoped for decades that speech synthesis based on articulatory geometry and dynamics would result in a breakthrough in quality and naturalness of speech synthesis, but this has not happened. It is now possible to generate high quality synthetic speech, such as with the Klatt synthesizer, by modeling only the properties (spectral, etc.) of the output signal."

In parallel, corpus-based speech synthesis approaches (for a review see [10, 11]) became more and more successful in order to realize high quality speech synthesis. Thus, the development of articulatory speech synthesis as high-quality synthesis was slowing down. Nevertheless, remarkable progress has been made in the last decade of this century concerning the increase in quality of articulatory speech synthesis (e.g., [12, 13]). Moreover, concerning a further development of articulatory speech synthesis systems as a research tool some interesting research questions could be focused on: (i) Can articulatory-acoustic models – which are used as front-end modules of neurobiologically motivated control models – contribute to the underpinning of neurobiologically oriented theories of speech and language acquisition and of speech and language processing (see e.g., [14-16])? (ii) Can articulatory-acoustic models be used profitably in medical research in order to investigate voice, speech, and language disorders (see e.g., [17, 18])? (iii) Can articulatory-acoustic models be used in phonetics and in foreign language

teaching in order to underpin theories of articulation and co-articulation in different languages (e.g., [19-21])? (iv) Does the refinement of self-oscillating vocal fold models and the refinement of aeroacoustic vocal tract models contribute significantly to the growth of physical-biological knowledge concerning the biomechanics and aerodynamics of the vocal tract (see e.g., [22, 23])? The above cited literature, which is already given here in the context of each question, already allows to answer all these questions positively.

2 Articulatory Models and Control Approaches for these Models

Articulatory speech synthesis models the "acustogenesis" (in German: "Akustogenese", a term introduced by Georg Heike [24]) or vocal tract acoustics, i.e., the mechanic-vibrational, acoustic, and aerodynamic processes taking place in the human vocal tract that cause the generation of speech sounds. In the pulmonary system, this is the generation of air pressure and air flow; in the laryngeal system, this is the generation of the vocal fold vibration and consequently the generation of the primary sound signal (phonation); and in the supralaryngeal system (vocal tract tube), this is the transformation of the primary sound by the resonances occurring in the vocal tract tube, including the loss mechanisms that occur there (e.g., dissipation through movement of the air molecules), as well as by the radiation of the sound from mouth and nostrils (e.g., [25], p. 103ff). In addition, the acustogenesis or vocal tract acoustics includes the generation of noise at further (often called: secondary) sound sources, i.e., friction noise, which is caused by a vocal tube constriction.

The simulation of these processes initially presupposes the modeling of the speech organs (larynx, tongue, soft palate, lips, hard palate, lower jaw, nasal cavities) that determine the shape of the vocal tube and their changes over time. This shaping of the vocal tube is calculated by articulatory models. In the simplest case, the form of the vocal tube and its change over time can be parameterized by using phonetically motivated parameters (phonetic control parameters) on the basis of X-ray or MRT data of a language-specific collection of speech sounds leading to *geometric articulation models* (two-dimensional models: e.g., [3, 26]; three-dimensional models: e.g., [27]). If the control parameters are obtained using statistical methods (e.g., using statistical factor analysis) based on a sequence of sound specific imaging data or on imaging data obtained from time-sequencies of continuously uttered speech, these basically geometrical models can also be labeled as *statistically based articulation models* (e.g., [4, 28]).

In the case of *biomechanically based articulatory models* the shape and position as well as the change in shape and position of the model articulators is determined solely from the biomechanical data of the articulators (e.g., mass, elasticity, compressibility). Here, the articulators are driven by patterns of neural activation of all individual muscles and muscle groups appearing in the vocal tract. A typical approach used here is the finite element method (two-dimensional models: e.g., [29-31]; three-dimensional models: e.g., [32, 33]). These models allow the control of speech articulation by means of neurobiologically based strategies. Thus, these approaches are able in addition to simulate the overall sequence of neural activation patterns at cognitive levels (i.e., modeling of concepts or intentions for communication, modeling of word selection for utterance formation, and modeling of articulatory motor processes including the activation of all speech articulator muscles involved in the realization of an utterance (e.g., [29, 14]). In addition to these neurobiologically detailed control models, *hierarchical state feedback models* have been developed successfully (e.g., [15]), which contain cognitively motivated approaches as well as sensorimotor approaches to control of articulatory models. Since biomechanical-neuromuscular articulatory models are still in an early stage of development, the front-end module used in all currently implemented neural based or cognitive-sensorimotor models are (easily controllable) geometrical or statistical articulation model (for a review of articulatory models see [34]).

3 Sets of Control Parameters and Basic Articulatory Production Units

In order to implement a quantitative approach for the control of articulation, two main questions have to be answered. Firstly: Are there natural and effective, cognitive-linguistic and/or neurobiologically motivated *sets of articulatory control parameters* of suitable (i.e., not too high) dimensionality? If we look at all geometric and statistical articulatory models, then mostly around 10 control parameters are assumed, e.g., degree of mouth opening, degree of lip rounding, height and palatal to velar position of the tongue dorsum, height and dental to postalveolar position of the tip of the tongue, degree of raising/lowering of the velum (soft palate), larynx position (height), glottal aperture, vocal cord tension, and lung pressure). All these parameters can also be directly interpreted phonetically [34].

In the case of biomechanical models, an amount of about 20-40 muscles or muscle groups need to be activated, which, however, interact synergistically with regard to the degrees of freedom of movements of the speech tract articulators, so that here the number of effective controls parameter lies between 10 and 20 [34]. A detailed neuromuscular model for controlling biomechanical articulatory models is the λ-model [29, 30, 33, 35]. This model implicates that the activation of a neuromuscular unit (i.e., of a muscle or a muscle group) is not only controlled top-down, but also depends on current values of state variables of the muscle itself (bottom-up control). In this model, the target deflection (i.e., target muscle length) λ of each muscle is used as the top-down control parameter for each muscle. Thus, the set of all target deflections of all muscles controlling an articulator describe the desired new target position of this articulator at the specific point in time. Due to the redundancy of the muscle system of each articulator, several sets of target deflections of the muscle system of an articulator (sets of λ values on an articulator) can lead to the same target displacement for that articulator. Thus, the λ-model implicates that target displacements play an important role as control variable even in this neurobiologically motivated control approach. The λ-model mainly describes how high-level control parameters which here represent displacement targets of articulators can be converted into a set of muscle controlling λ-values as a function of time. Thus, this model of neuromuscular control also implicates that when reducing the neurobiological level of detail of a control approach, it makes sense to use geometrically based articulation model directly controlled by displacement-depending and articulator-related parameters (see e.g., the DIVA/GODIVA model: [14, 36]; or the FACTS model: [15]).

The second question is: What are the basic cognitive-linguistic or phonetic *units of articulatory control*? Since most speech synthesis systems are controlled by a chain of speech sounds plus suprasegmental information, the answer could be: The basic units of control are syllables, demi-syllables, or speech sounds, i.e., *segments*. Such a segmental control approach for articulatory speech synthesis including a concept of articulatory-segmental underspecification has been suggested by Kröger [37]. In recent decades, however, the concept of *articulatory gestures*, also known as articulatory movement units or as *dynamic movement primitives*, got remarkable attention [19, 21, 38, 39]. Like a chain of phonemes, articulatory gestures can also describe a syllable, word, or phrase on the phonological as well as on the articulatory level. But in contrast to a phoneme chain, gestures are ordered on parallel tiers in time and here, co-articulation (as well as several segmental reduction phenomena) directly result from the temporal overlapping of articulatory gestures (co-production of gestures; see simulation results using articulatory models and gestural control approaches, e.g., [20, 34, 39, 40]).

4 Vocal Tract Acoustics

Acoustogenesis, i.e., the generation of the acoustic speech signal in the vocal tract tube, includes not only acoustic but also mechanical and aerodynamic aspects. Thus acoustogenesis comprises (i) modeling of the respiratory system to realize temporal change of lung volume for simulating subglottal pressure and glottal airflow during speech, (ii) modeling of the vibration behavior of the vocal folds for simulating primary sound generation (phonation), (iii) simulation of sound wave propagation and of air flow in the vocal tract tube, (iv) simulation of fricative sound generation at secondary sound sources (temporarily occurring vocal tube constrictions), and (v) simulation the speech sound propagation from mouth and nostrils. In the case of primary and secondary sound sources, the modeling of acoustic as well as aerodynamic aspects is needed in order simulate turbulent noise generation correctly. The strength of the airflow, its change over time due to changes in the cross-sectional area of constrictions and due to changes in air pressure upstream of each constriction as well as the strength and the acoustic consequences of turbulences appearing downstream of each constriction in the vocal tract tube need to be considered in the modelling approaches.

An early computer-implemented model that already takes into account all of the above acoustic and aerodynamic mechanisms is the speech synthesis model published by Liljencrants [5]. This is an extension of the reflection-type line analog published by Kelly & Lochbaum [41] by including all important loss mechanisms occurring in the vocal tract. However, the Liljencrants model did not include a self-oscillating vocal fold model. The vibration pattern of the vocal folds is imprinted here in the synthesizer. But Liljencrants modeled and implemented all acoustically and aerodynamically important loss mechanisms appearing in the vocal tract tube which led to an suitable approximation of frequency-dependent damping of the acoustic speech signal as well as to the estimation of the DC component of the volume flow within the vocal tract tube. The modelled losses are serial and parallel losses from the viewpoint

of electrical circuit representation of the vocal tract tube (see e.g., [25], p. 103ff). Moreover, the Liljencrants model can easily be combined with self-oscillating vocal fold models (e.g., [25], p. 92ff), whereby the forces acting on the vocal folds are derived from aerodynamic parameters (air pressure and amount of air flow occurring in, directly below, and directly above the glottal constriction).

A major disadvantage of this approach, however, is that the resonance behavior of the vocal tract tube appearing above the third formant cannot be simulated correctly. In order to eliminate this disadvantage, a hybrid time-frequency domain model (although computationally time-consuming) for the simulation of acoustogenesis was developed [42]. A further and perhaps more serious disadvantage of the simple reflection-type line model is the difficulty in modeling changes in the length of the vocal tract tube, such as those that occur during the transition from the vowel [i] or [a] to [u]. Thus, even in pure time domain models, the simple reflection-type line analog of Kelly & Lochbaum [41] is no longer used today, but instead, for example, the time domain approach of Birkholz & Jackèl [43] which is further developed in the last years allows to overcome these problems.

Another disadvantage of the abovementioned simple reflection-type line analog is that it only models sound propagation in one dimension, namely along the center line of the vocal tube, i.e., along the path through the tracheal, laryngeal, pharyngeal and oral regions (from the lungs through the trachea, glottis, pharynx and mouth to the lips), extended by a branch into the nasal cavity. In addition to this one-dimensional approach, there are now far more complex approaches both for modeling the vocal fold oscillation behavior for generating the phonatory sound signal (e.g., [44-47]) and for modeling the sound propagation in the vocal tube in combination with the secondary sound generation [48-56].

Since it is difficult to simulate the acoustics and aerodynamics especially in the case of noise generation by turbulent flow downstream a potential vocal tube constriction (glottal and supraglottal), in addition mechanical-aeroacoustic models and mathematical models for the investigation and description of all aeroacoustic phenomena occurring at primary and secondary sound sources have been developed [57-63]. From this, new simulation approaches have been developed which now take into account more detailed knowledge of aeroacoustic phenomena appearing at the secondary sound sources (e.g., [60]) or at the primary sound source (e.g., [23, 64]).

5 Simulation of Symptoms of Voice, Speech, and Language Disorders

There are already several complex vocal fold models that are able to simulate symptoms of different types of voice disorders. Falk et al. [17] presents an aeroacoustic model in which the vibration of the vocal folds is externally specified. But this model is able to evaluate the energy transfer from the aerodynamic towards the mechanical system of the vocal folds for different types of vocal fold vibration patterns. It is shown that the efficiency of this energy transfer decreases with decreasing duration of glottal closure and with decreasing maximal contact area during glottal closure. It is shown that this reduced energy transfer especially leads to a loss of energy in higher spectral regions and in the overall strength of the generated phonatory sound signal. Moreover, it has been shown using this modeling approach that the asymmetry of vocal fold vibration leads to a decrease in the intensity of the phonatory sound signal. Thus, these simulations underline the fact, that a glottal closing insufficiency as well as a left-right vibration asymmetry of the vocal folds can severely impair the acoustic-perceptual quality of the phonatory sound signal.

Zangh & Jiang [65] modeled vocal fold vibrations in a vocal fold model including a vocal fold polyp attached to one side of the vocal folds. The basic vocal fold model here is a self-oscillating two-mass model and the polyp is added here as a further small mass placed on one of both vocal folds. This study shows that a vocal fold polyp will negatively influence the duration and type of glottal closure (complete or incomplete) and in particular the addition of this extra mass can lead to aperiodic vocal fold vibrations. It was also shown that the modeled physical parameters of the polyp (size, stiffness and damping) have differently strong effects on the resulting vocal fold vibration. In the case of a very large polyp, even subharmonic vibration patterns and chaotic vibration behavior of the vocal folds appear.

Tao & Jiang [66] present a complex finite element model of the vocal folds. This model is able to estimate the mechanical (over-)loads appearing in specific temporal portions of the glottal cycle as well as at different places of the vocal folds. This allows predictions to be made as to the extent to which certain aerodynamic and geometric preconditions lead to particularly high mechanical loads at various points within and on the surface of the vocal folds. This model thus allows predictions regarding the onset of voice fatigue and enables predictions of potential injuries to the vocal folds due to overuse.

Language and speech disorders comprise disorders in the cognitive-linguistic subsystem of the speech processing system (the aphasias), disorders in the sensorimotor subsystem of the speech processing system (apraxia of speech and the dysarthrias) and anatomical-functional disorders in the articulation apparatus (articulation disorders). Language disorders can cause errors in the generation of the sound sequence of an utterance, i.e., errors appearing at the output level of the cognitive-linguistic subsystem of speech processing (i.e., at the phonological level). These errors can be simulated using neurobiologically based control systems of articulatory speech synthesis (e.g., [34, 67]). Thus, the simulation of language disorders also requires the modeling of the entire neural network of speech production, while the modeling of the motor and articulatory-acoustic realization of an utterance is not required here. Roelofs [67] and Kröger [68] describe the simulation of typical symptoms of various forms of aphasia using such neurobiologically inspired production models (neurologically inspired neural network models).

Speech disorders include anatomical-functional disorders at the level of the speech apparatus (articulation disorders), disorders at the level of neuromuscular activation and its control (some subtypes of dysarthria, see e.g., [18]) and disorders at the level of motor planning (apraxia of speech, see e.g., [69]). From the viewpoint of modeling of speech production, a model-theoretical classification of different forms of dysarthria and a model-theoretical differentiation of apraxia of speech with respect to different neurogenic causes is suggested [18, 70, 71]. A model-theoretical classification of articulation disorders is given, for example, in [72] from the point of view of articulatory phonology. Stuttering can also be labeled as language and speech disorder and causes disturbances in the flow of articulation. Symptoms of neurogenic stuttering can be simulated using a neural control model comprising cognitive-linguistic as well as sensorimotor aspects [73]. A more extensive etiological description of stuttering from a model-oriented point of view is given in [74]. Thus, neurobiologically oriented production models, even if designed for the control of an articulatory-aeroacoustic synthesizer as a front-end module, can contribute to the simulation of typical symptoms of these speech disorders and thus help to understand the etiology of these disorders.

6 Discussion

On the one hand, the development of articulatory speech synthesis can continue to be pursued with the aim of achieving high acoustic signal quality even on the basis of not too complex parametric articulatory model approaches. These type of models (e.g., Birkholz [12]) already have sufficient flexibility for articulatory and acoustic optimization (e.g., Krug et al. [13]). An increase in the level of detail that could be achieved using biomechanical models as control modules for articulatory models is not necessarily promising with respect to the complexity of neuromuscular control if high speech signal quality is aimed for. In case of these neurobiologically and biomechanically based approaches a simplification of the control model as well as of the used articulatory model is needed by means of optimizing and minimizing the set of high-level control parameters (see the approaches suggested by Sanguineti et al. [29], Perrier et al. [30], and Buchaillard et al. [33]).

On the other hand, there is a potential field of application for articulatory speech synthesis in its use as a front-end module for research-oriented neurobiologically based production models. These approaches for controlling articulatory speech synthesizers can make a valuable contribution to solving fundamentally questions about the neuronal foundation of articulatory control: How can an effective set of “higher-level” articulatory control parameters be derived in case of neuromuscular based control? Which are the basic temporal-spatial units of movement in speech production? How relevant are control units like segments (phonemes), distinctive features, or articulatory gestures in cognitive-sensorimotor models of speech production? To answer these questions, the realization of the articulatory speech synthesis system including a neurobiological control module (neural cognitive-sensorimotor model of speech processing, see e.g., [16]) plus neuromuscular detailed control concepts (e.g., [29]) and plus a biomechanical and aeroacoustic detailed realization at the level of the articulatory and aeroacoustic model is required.

Articulatory synthesis - especially in connection with the cognitive-sensorimotor production models implemented on top of the articulatory model - can also be used as a systematic tool to explore and describe voice, speech, and language disorders. Since both the cognitive-sensorimotor neural system and the neuromuscular-articulatory system cannot be observed in all its variables during natural speech production processes, the observations derived from model simulations (“simulated data” like e.g., the time-dependent changes in neuronal activation at different levels of the language and speech system)

and the observation of the resulting (simulated) articulatory movements and of all aerodynamic and acoustic parameters coming with these simulations gives important information, especially because neural or biomechanical dysfunctions can be introduced ("inserted") into these models in a controlled way in order to simulate specific speech or language disorders in an absolute well-defined way (diagnosis of patients could be broad and the exact neural cause of a disorder diagnosed for a patient is not easily and exact determinable). However, there is still no comprehensive neurobiological production model comprising all sensorimotor aspects (current approaches are e.g., Guenther [14] and Bohland et al. [36]) or al cognitive linguistic aspects (current approaches are e.g., Roelofs [67] or Kröger et al. [16]) combined with a biomechanical articulation model for realizing the muscular control of individual articulators (as e.g., in [29]) and in combination with an realistic articulatory-acoustic synthesis model (as e.g., proposed by [12, 30]). Such a comprehensive production and synthesis model is still a challenging research endeavor but if developed it could serve as fruitful research tool in the fields of medicine, psychology, linguistics, as well as in speech acoustics.

7 Literature

[1] VON KEMPELEN, W. (1791) Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine. Stuttgart, Germany, Frommann-Holzboog.

[2] ISHIZAKA, K., FLANAGAN, J.L. (1972) *Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords.* Bell Systems Technical Journal 51, 1233-1268.

[3] MERMELSTEIN, P. (1973) *Articulatory Model for the Study of Speech Production.* Journal of the Acoustical Society of America 53, 1070-1082.

[4] MAEDA, S. (1979) *An Articulatory Model of the Tongue Based on a Statistical Analysis*. Journal of the Acoustical Society of America 65, S22.

[5] LILJENCRANTS, J. (1985) *Speech synthesis with a reflection-type line analog.* Proceedings of the Royal Institute of Technology, Vol. 85, No. 2. Stockholm, Sweden, RIT Press.

[6] COKER, C.H. (1976) *A Model of Articulatory Dynamics and Control.* Proceedings of IEEE 64, 452-460.

[7] 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.

[8] 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.

[9] WILHELMS-TRICARICO, R.F., PERKELL, J.S. (1997) *Biomechanical and Physiologically Based Speech Modeling.* In: J.P.H. van Santen, R.W. Sproat, J.P. Olive, and J. Hirschberg (eds.) *Progress in Speech Synthesis*, New York, USA, Springer-Verlag, pp. 221-234.

[10] ZEN, H., TOKUDA, K., AND BLACK, A.W. (2009) *Statistical Parametric Speech Synthesis*. Speech Communication 51, 1039-1064.

[11] KAHN, R.A., CHITODE, J.S. (2016) *Concatenative Speech Synthesis: A Review*. International Journal of Computer Applications 136, 1-6.

[12] BIRKHOLZ, P. (2013) Modeling consonant-vowel coarticulation for articulatory speech synthesis. PLoS ONE, 8:e60603.

[13] KRUG, P., STONE, S., BIRKHOLZ. P. (2021) Intelligibility and naturalness of articulatory synthesis with VocalTractLab compared to established speech synthesis technologies. Proceedings of 11th ISCA Speech Synthesis Workshop 2021. Budapest, Hungary, pp. 102-107.

[14] GUENTHER, F.H. (2006) *Cortical Interactions Underlying the Production of Speech Sounds.* Journal of Communication Disorders 39, 350-365.

[15] PARRELL, B., RAMANARAYANAN, V., NAGARAJAN, S., HOUDE, J. (2019) *The FACTS model of speech motor control: Fusing state estimation and task-based control.* PLoS Computational Biology 15:e1007321.

[16] KRÖGER, B.J., BEKOLAY, T., CAO, M. (2022). On the Emergence of Phonological Knowledge and on Motor Planning and Motor Programming in a Developmental Model of Speech Production. Frontiers in Human Neuroscience 16:844529.

[17] FALK, S., KNIESBURGES, S., SCHODER, S., JAKUBAß, B., MAURERLEHNER, P., ECHTERNACH, M., KALTENBACHER, M., DÖLLINGER, M. (2021) *3D-FV-FE Aeroacoustic Larynx Model for Investigation of Functional Based Voice Disorders.* Frontiers in Physiology 12:616985.

[18] KEARNEY, E., GUENTHER F.H. (2019) *Articulating: the neural mechanisms of speech production.* Language, Cognition and Neuroscience 34, 1214-1229.

[19] BROWMAN, C.P., GOLDSTEIN, L. (1992) *Articulatory phonology: An overview.* Phonetica 49, 155-180.

[20] KRÖGER, B.J. (1993) A gestural production model and its application to reduction in German. Phonetica 50, 213-233.

[21] HALL, N. (2010) *Articulatory phonology.* Language and Linguistics Compass 4, 818-830.

[22] YOSHINAGA, T., NOZAKI, K., WADA, S. (2019) *Aeroacoustic analysis on individual characteristics in sibilant fricative production.* Journal of the Acoustical Society of America 146, 1239-1251.

[23] SCHODER, S., WEITZ, M., MAURERLEHNER, P., HAUSER, A., FALK, S., KNIESBURGES, S., DÖLLINGER, M., KALTENBACHER, M. (2020) *Hybrid Aeroacoustic Approach for the Efficient Numerical Simulation of Human Phonation.* Journal of the Acoustical Society of America 147, 1179-1194.
[24] HEIKE, G, THÜRMANN, E. (1980) *Phonetik*. In: Althaus, H.P., Henne, H., Wiegand, H.E. (Eds.): *Lexikon der Germanistischen Linguistik*. Berlin, New York, Max Niemeyer Verlag, pp. 120-128.
[25] KRÖGER, B.J. (1998) *Ein phonetisches Modell der Sprachproduktion*. Tübingen, Germany, Niemeyer-Verlag.
[26] ISKAROUS, K., GOLDSTEIN, L., WHALEN, D., TIEDE, M., RUBIN, P. (2003) *CASY: The Haskins Configurable Articulatory Synthesizer*. Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, Spain, pp. 185-188.
[27] BIRKHOLZ, P., JACKÈL, D. (2003) *A Three-Dimensional Model of the Vocal Tract for Speech Synthesis*. Proceedings of the 15th International Congress of Phonetic Sciences - ICPhS'2003. Barcelona, Spain, pp. 2597-2600.
[28] ENGWALL, O. (2003) Combining MRI, EMA and EPG Measurements in a Three- Dimensional Tongue Model. Speech Communication 41, 303-329.
[29] SANGUINETI, V., LABOISSIÈRE, R., OSTRY, D. J. (1998) *A Dynamic Biomechanical Model for Neural Control of Speech Production*. Journal of the Acoustical Society of America 103, 1615-1627.
[30] PERRIER, P., PAYAN, Y., ZANDIPOUR, M., PERKELL, J. (2003) *Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study*. Journal of the Acoustical Society of America 114, 1582-1599.
[31] DANG J., HONDA, K. (2004) *Construction and control of a physiological articulatory model*. Journal of the Acoustical Society of America 115, 853-870.
[32] WILHELMS-TRICARICO, R. (1996) A Biomechanical and Physiologically-Based Vocal Tract Model and its Control. Journal of Phonetics 24, 23-38.
[33] BUCHAILLARD, S., PERRIER, P., PAYAN, Y. (2009) A Biomechanical Model of Cardinal Vowel Production: Muscle Activations and the Impact of Gravity on Tongue Positioning. Journal of the Acoustical Society of America 126, 2033-2051.
[34] KRÖGER, B.J. (2022) Computer-Implemented Articulatory Models for Speech Production: A Review. Frontiers in Robotics and AI 9:796739.
[35] LABOISSIÈRE, R., OSTRY, D.J., FELDMAN, A.G. (1996) *The control of multi-muscle systems: human jaw and hyoid movements*. Biological Cybernetics 74, 373-384.
[36] BOHLAND, J.W., BULLOCK, D., GUENTHER, F.H. (2010) *Neural Representations and Mechanisms for the Performance of Simple Speech Sequences*. Journal of Cognitive Neuroscience 22, 1504-1529.
[37] KRÖGER, B.J. (1992) *Minimal rules for articulatory speech synthesis*. In: J. Vandewalle, R. Boite, M. Moonen, A. Oosterlinck (eds.) *Signal Processing VI: Theories and Applications.* Amsterdam, NL, Elsevier, pp. 331-334.
[38] KRÖGER, B.J. (2018) Neuronale Modellierung der Sprachverarbeitung und des Sprachlernens. Eine Einführung. Springer Verlag, Berlin, Heidelberg, New York.
[39] PARRELL, B., LAMMERT, A.C. (2019) Bridging Dynamical Systems and Optimal Trajectory Approaches to Speech Motor Control With Dynamic Movement Primitives. Frontiers in Psychology 10:3389.
[40] KRÖGER, B.J., BIRKHOLZ, P. (2007) 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, pp. 174-189.
[41] KELLY, J.L., LOCHBAUM, C.C. (1962) *Speech synthesis.* Proceedings of the International Congress on Acoustics, Paper G-42, S. 1-4. Reprinted in: J.L. Flanagan, L.R. Rabiner (eds.), *Speech Synthesis.* Stoudsburg, USA, Dowden, Hutchinson & Ross, pp. 127-130.
[42] 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.
[43] BIRKHOLZ P., JACKÈL D. (2004) *Simulation of flow and acoustics in the vocal tract*. Proceedings of 30th Deutsche Jahrestagung für Akustik (CFA/DAGA 2004). Strasbourg, France, pp. 895-896
[44] PELORSON, X., HIRSCHBERG, A., VAN HASSEL, R. R., WIJNANDS, A. P. J., AUREGAN, Y. (1994) Theoretical and experimental study of quasisteady- flow separation within the glottis during phonation - application to a modified two-mass model. Journal of the Acoustical Society of America 96, 3416-3431.
[45] NARAYANAN, S., ALWAN, A. (2000) *Noise source models for fricative consonants*. IEEE Transactions on Speech and Audio Processing 8, 328-344.
[46] YANG, A., LOHSCHELLER, J., BERRY, D. A., BECKER, S., EYSHOLDT, U., VOIGT, D., DÖLLINGER, M. (2010) *Biomechanical modeling of the three-dimensional aspects of human vocal fold dynamics*. The Journal of the Acoustical Society of America 127, 1014–1031.
[47] BIRKHOLZ, P., DRECHSEL, S., STONE, S. (2019) *Perceptual optimization of an enhanced geometric vocal fold model for articulatory speech synthesis*. Proceedings Interspeech 2019. Graz, Austria, pp. 3765-3769.
[48] SINDER, D.J., KRANE, M.H., FLANAGAN, J.L. (1998) *Synthesis of fricative sounds using an aeroacoustic noise generation model*. The Journal of the Acoustical Society of America 103, 2775.
[49] HUANG, J., LEVINSON, S., DAVIS, D., SLIMON, S. (2002) *Articulatory Speech Synthesis Based upon Fluid Dynamic Principles*. Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, Florida, USA, pp. 445-448.

[50] BIRKHOLZ, P., JACKÈL, D., KRÖGER, B.J. (2007) *Simulation of losses due to turbulence in the time-varying vocal system*. IEEE Transactions on Audio, Speech, and Language Processing 15, 1218-1225.
[51] TAKEMOTO, H., MOKHTARI, P., KITAMURA, T. (2010) *Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method*. Journal of the Acoustical Society of America 128, 3724-3738.
[52] LEVINSON, S., DAVIS, D., SLIMON, S., HUANG, J. (2012) *Articulatory Speech Synthesis from the Fluid Dynamics of the Vocal Apparatus*. San Rafael, CA, USA, Morgan & Claypool.
[53] ARNELA, M., GUASCH, O. (2014) *Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels*. The Journal of the Acoustical Society of America 135, 369-379.
[54] ZAPPI, V., VASUDEVAN, A., FELS, S. (2016) *Towards real-time two-dimensional wave propagation for articulatory speech synthesis*. The Journal of the Acoustical Society of America 139, p. 2010.
[55] PONT, A., GUASCH, O., BAIGES, J., CODINA, R., VAN HIRTUM, A. (2018) *Computational aeroacoustics to identify sound sources in the generation of sibilant /s/*. International Journal for Numerical Methods in Biomedical Engineering 35:e3153.
[56] BLANDIN, R., ARNELA, M., FÉLIX, S., DOC, J.B., BIRKHOLZ, P. (2022) Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements. IEEE Access 10, 69922-69938.
[57] MCGOWAN, R.S. (1988) *An aeroacoustic approach to phonation*. Journal of the Acoustical Society of America 83, 696-704.
[58] PELORSON, X., HOFMANS, G.C.J., RANUCCI, M., BOSCH, R.C.M. (1997) *On the fluid mechanics of bilabial plosives*. Speech Communication 22, 155-172.
[59] KRANE, M.H. (2005) *Aeroacoustic production of low-frequency unvoiced speech sounds.* The Journal of the Acoustical Society of America 118, 410-427.
[60] HOWE, M.S., MCGOWAN, R.S. (2005) *Aeroacoustics of [s]*. Proceedings of the Royal Society A, 461, 1005-1028.
[61] MCGOWAN, R.S., HOWE, M.S. (2012) *Source-tract interaction with prescribed vocal fold motion*. The Journal of the Acoustical Society of America 131, 2999-3016.
[62] MCPHAIL, M.J., CAMPO, E.T., KRANE, M.H. (2019) *Aeroacoustic source characterization in a physical model of phonation*. The Journal of the Acoustical Society of America 146, 1230-1238
[63] MOTIE-SHIRAZI, M., ZANARTU, M., PETERSON, S.D., ERATH, B.D. (2021) *Vocal fold dynamics in a synthetic self-oscillating model: Intraglottal aerodynamic pressure and energy*. The Journal of the Acoustical Society of America 150, 1332-1345.
[64] SCHICKHOFER, L., MIHAESCU, M. (2020) Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes. Journal of Biomechanics 99: 109484.
[65] ZHANG, Y., JIANG, J.J. (2004) *Chaotic vibrations of a vocal fold model with a unilateral polyp*. Journal of the Acoustical Society of America 115, 1266-1269.
[66] TAO, C., JIANG, J.J. (2007) Mechanical stress during phonation in a self-oscillating finite-element vocal fold model. Journal of Biomechanics 40, 2191-2198.
[67] ROELOFS, A. (2014) A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. Cortex 59, 33-48.
[68] KRÖGER, B.J., STILLE, C.M., BLOUW, P., BEKOLAY, T., STEWART, T.C. (2020) Hierarchical Sequencing and Feedforward and Feedback Control Mechanisms in Speech Production: A Preliminary Approach for Modeling Normal and Disordered Speech. Frontiers in Computational Neuroscience 14:573554.
[69] VAN DER MERWE, A. (2021) New perspectives on speech motor planning and programming in the context of the four-level model and its implications for understanding the pathophysiology underlying apraxia of speech and other motor speech disorders. Aphasiology 35, 397-423.
[70] MILLER, H.E., GUENTHER, F.H. (2021) Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework. Aphasiology 35, 424-441.
[71] KRÖGER, B.J. (2021) Modeling dysfunctions in the coordination of voice and supraglottal articulation in neurogenic speech disorders. In: C. Manfredi (ed.) Models and Analysis of Vocal Emissions for Biomedical Applications. Firenze, Italy, Firenze University Press, pp. 79-82.
[72] NAMASIVAYAM, A.K., COLEMAN, D., O'DWYER, A., VAN LIESHOUT, P. (2020) *Speech Sound Disorders in Children: An Articulatory Phonology Perspective*. Frontiers in Psychology 10:2998.
[73] CIVIER, O., BULLOCK, D., MAX, L., GUENTHER, F.H. (2013) Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. Brain and Language 126, 263-278.
[74] CHANG, S.E., GUENTHER, F.H. (2020) Involvement of the Cortico-Basal Ganglia-Thalamocortical Loop in Developmental Stuttering. Frontiers in Psychology 10:3088.