MODELING DYSFUNCTIONS IN THE COORDINATION OF VOICE AND SUPRAGLOTTAL ARTICULATION IN NEUROGENIC SPEECH DISORDERS

Bernd J. Kröger

Department of Phoniatrics Pedaudiology and Communication Disorders,
RWTH Aachen University, Aachen, Germany

Abstract:
Neurogenic speech disorders like apraxia of speech or dysarthria show various symptoms in speech and vocalizations. Many of these symptoms can be simulated using a neural model of speech production which includes components for linguistic planning, motor planning, motor programming, and articulatory execution. The execution module (articulatory-acoustic synthesizer) comprises a supra-laryngeal, laryngeal, and sub-laryngeal part and generates normal as well as disordered vocal fold and articulator movements and it generates normal as well as disordered phonation and speech signals.

The concept of gestures as target-directed articulator movements (gestures) is of central importance in our approach. In this paper we concentrate on the simulation of dyscoordinating and of over- and undershooting articulatory and phonatory gestures. The resulting simulated acoustic signals will be compared to natural acoustic signals of normal and disordered speech and vocalizations.

Keywords: **Neurogenic speech disorders, speech gestures, coordination of gestures, articulatory overshoot, articulatory undershoot**

I. INTRODUCTION

Apraxia of speech is defined as a deficit in planning of speech while *dysarthia* is defined as a deficit in motor programming and neuromuscular execution [1]. Both types of speech disorders affect the control of the supra-laryngeal as well as for the laryngeal and sub-laryngeal domain (articulation, phonation, respiration) and these speech disorders affect the segmental level, i.e., lead to distortions of speech sounds, as well as to a distortion of intonation and of syllabic stress patterns. Deficits in speech motor (apraxia of speech) result in deficits in temporal coordination of gestures within and between all three domains (articulation, phonation, respiration) as well as in deficits in correct implementation of the movement target for each single gesture. Planning deficits are mainly due to neural dysfunctions in premotor areas and motor cortex. Deficits in speech motor programming and execution (dysarthria) affect the realization of each gesture by distortions in gesture control and in gesture execution. That leads to imprecise realizations of gestures with respect to gesture duration but mainly with respect to target reaching. Programming, execution, and control deficits are mainly due to neural dysfunctions in motor neurons, basal ganglia, and/or cerebellum.

In this paper we will concentrate on selected symptoms of different speech disorders. Patients suffering from *apraxia of speech* (planning deficits resulting from dysfunctions at different cortical locations) show symptoms like groping, speech sound distortions, articulation errors in producing complex syllables, slow speech rate, and syllable segregation [1]. In case of dysarthria, we need to separate different subtypes [2, 3]. Patients suffering from *ataxic dysartria* (control deficits resulting from cerebellar dysfunctions) show slow and irregular articulatory movement rates and high variability in syllable intensity level. Patient suffering from *flaccid dysarthria* (lower motor neuron damage) show symptoms like breathy voice, short phrases, increased nasal resonance resulting from imperfect closure of the velopharyngeal port and imprecise articulation. *Spastic dysarthria* (bilateral damage of upper motor neurons) leads to symptoms like strained voice and slow articulation resulting from too high muscle tonus. *Hypokinetic dysarthria* (control deficit resulting from basal ganglia dysfunctions) leads to low movement amplitudes while *hyperkinetic dysarthria* (same) leads to involuntary strong and imprecise movements, which not necessarily result from high articulatory effort.

The concept of *speech gestures* [4, 5] allows to explain the speech and voicing symptoms mentioned above by checking the temporal coordination of gestures within a syllable as well as by introducing the idea of gesture target overshoot and gesture target undershoot. Gestures can be defined for the supra-laryngeal system (*vocalic gestures, consonantal gestures,* and *velopharyngeal gestures*, Kröger & Birkholz 2007, p. 181ff) as well as for the laryngeal and sub-laryngeal system. In case of the laryngeal (glottal) system we can differentiate *glottal gestures* controlling vocal fold tension and glottal gestures controlling the positioning of the arytenoids. The later are glottal opening gestures for

producing unvoiced speech sounds, glottal closing gestures for producing phonation, and glottal tight closing gestures for producing glottal stop sounds (ibid.). In the case of the sub-laryngeal system, *pulmonary gestures* can be defined. The goal of these gestures is to control subglottal pressure as well as the time span for which a certain degree of subglottal pressure can be hold and for which a certain amount of airflow can be guaranteed to enable phonation as well as secondary sound source excitation.

Gestures always define *target-directed articulator movements*. The goal of each gesture is to reach an acoustically or perceptually relevant target state. In case of articulatory gestures, the target defines a spatial positioning of articulators within the vocal tract, e.g., for reaching vocalic tract shapes or for reaching consonantal constrictions or closures. In the case of glottal gestures, a target is defined as the positioning of the vocal folds or as a certain degree of vocal fold tension. In the case of pulmonary gestures, a target is defined dynamically as the dynamic change in lung volume which leads to the generation of a specific level of subglottal pressure.

II. METHODS

A. Description of the model

The model comprises a neural control component and an articulatory-acoustic model (Fig. 1). A compete linguistic description of the utterance, i.e., a narrow phonological transcription (linguistic input) is transformed into a gesture score (motor plan). The specification of the gesture score, i.e., the temporal coordination of all gestures of an utterance is called *motor planning* which takes place in the premotor area of the brain. The specification of each gesture with respect to the resulting neuromuscular activity is called *motor programming* and leads to a specific neural activation pattern for each syllable at the level of the primary motor cortex. The *execution* of gestures or motor programs is performed by the neuromuscular units of all articulators which lead to defined articulator movements for all articulators of all model components, i.e., for the movements controlling the pulmonary system (lung volume), for the movements controlling the vocal fold positioning, and for the movements controlling the lower jaw, tongue body, tongue tip, velum, and lips. A more detailed discussion concerning the separation of motor planning, motor programming and execution can be found in [1].

Movements of many articulators directly result from neuromuscular activations generated by the neural control component. But in the case of the vocal folds the control component only determines the (rest-)positioning of the folds for phonation or for producing voiceless sounds, while the vocal fold vibration patterns is initiated and controlled by aerodynamic states. The same holds for vocal tract articulators like lips, tongue tip and uvula in case of trills (/B/, /r/, and /R/).

While the brain locations for planning and activating motor programs are cortical, and while the execution of motor programs is mainly done via a direct feedforward motor neuron pathway, *somatosensory feedback signals,* i.e., *tactile and proprioceptive signals* are processed by the basal ganglia-thalamus complex as well as by the cerebellum for controlling and for eventually correcting motor programs and for altering motor programs and motor plans in case of articulatory distortions or in case of changes in the production system due to aging or disorders (feedback processing pathway in Fig. 1 and see [1]).

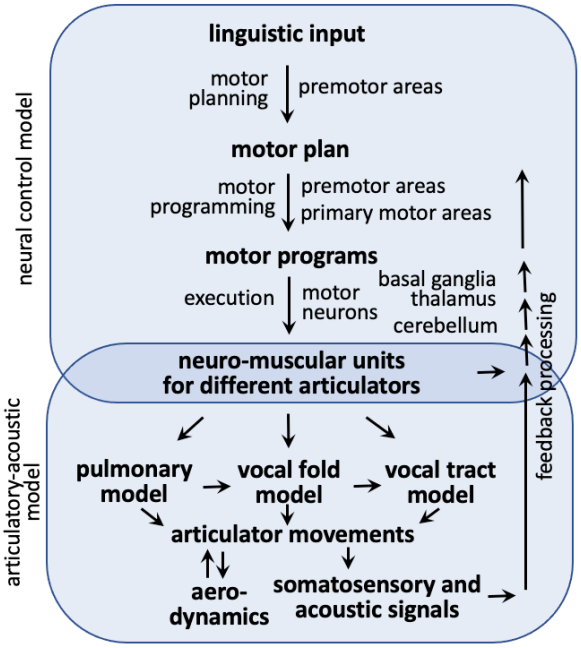

Figure 1: The production model

The articulatory-acoustic model comprises a pulmonary model for generating subglottal pressure and airflow, a self-oscillating vocal fold model for generating vocal fold oscillations for phonation and a vocal tract model for generating vocal tract shapes as function of time. The acoustic glottal source signal in modified in the vocal tract and is radiated it from the lips and nostrils [6, 7].

The motor plan of an utterance is specified as gesture score. A gesture score for the utterance or word (example: [pani]) is given in Fig. 2. The gestures are ordered in six tiers and the gesture targets are named for each gesture: (i) the targets of *vocalic gestures* describes the global form of the vocal tract (global tract form gestures: low tongue body -> /a/; high front location of tongue body -> /i/; high back location of tongue body -> /u/; the labial part of vocalic gestures, i.e., rounded or spread lips, is not displayed in Fig. 2); (ii) the targets of *consonantal gestures* describe the formation of a local

constriction or closure within the vocal tract (local tract constriction gestures: labial closing gestures -> /b/, /p/, /m/; apical closing gestures -> /d/, /t/, /n/, velar closing gestures -> /g/, /k/, /N/; apical near closing gestures -> /s/, /z/, etc.; see [4]).

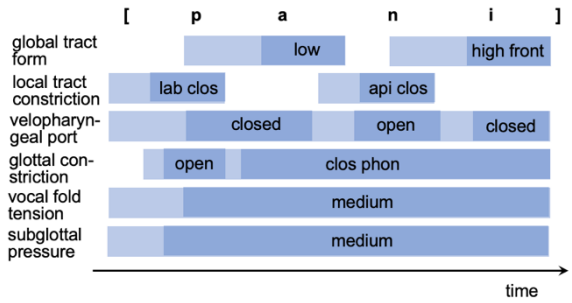

Figure 2: Gesture score of [pani]

(iii) *velopharyngeal opening vs. closing gestures* realize nasal vs. oral speech sounds (tight velopharyngeal closure in case of obstruents, i.e., in case of plosives and fricatives for guaranteeing a pressure built-up in the oral cavity during oral closure); (iv) *glottal opening vs. closing gestures* realize voiceless vs. voiced sounds (tight glottal closure to produce a glottal stop); (v) the target of a *vocal fold tension gesture* defines a F0-target within the intonation contour of an utterance (targets: low, medium, high tension of vocal folds); (vi) the target of a *pulmonary gesture* is holding a specific level of subglottal pressure over the whole time interval of an utterance (targets: low, medium, high in order to realize a soft, normal, or loud voice).

The light blue bars (including the dark blue portions) in Fig. 2 indicate the duration of activation for each gesture. The light blue time interval marks the *movement phase* of a gesture, while the dark blue time interval marks the period in which the gesture reached its target (*target phase*). In the case of vocalic tract-shaping gestures the movement phase is mainly hidden behind a local consonantal tract constriction. In the case of consonantal tract constriction gestures the movement phase occurs within the target phases of vowels and thus allows the perception of place of articulation by the appearance of audible formant transitions.

The gesture targets define (i) the main characteristics of the speech sounds like vocalic formant pattern (vocalic gestures), manner and place of articulation (consonantal gestures), nasal or oral realization of a speech sound (velopharyngeal gestures), voiced or unvoiced realization of a speech sound (glottal gestures), or they define (ii) important supraglottal features of an utterance like current F0-level (vocal fold tension gesture), current loudness or stress level (pulmonary gesture).

A *normal realization, an undershot, overshoot, and a corrected overshoot realization of a gesture* is shown in Fig. 3. Target over- or undershoot can be defined if gesture targets have spatial dimensions (vocalic tract shapes, consonantal closures or constrictions, degree of opening of velopharyngeal port or of glottal constriction) or if targets are defined in the acoustic or aerodynamic domain as frequency value or as pressure level. Target overshoot can be corrected during gesture execution by reversing the movement direction at a certain point in time (Fig. 3, bottom). In case of undershoot the duration of gesture activation interval (of gesture movement phase) needs to be extended or the articulator velocity must be increased (not shown in Fig. 2).

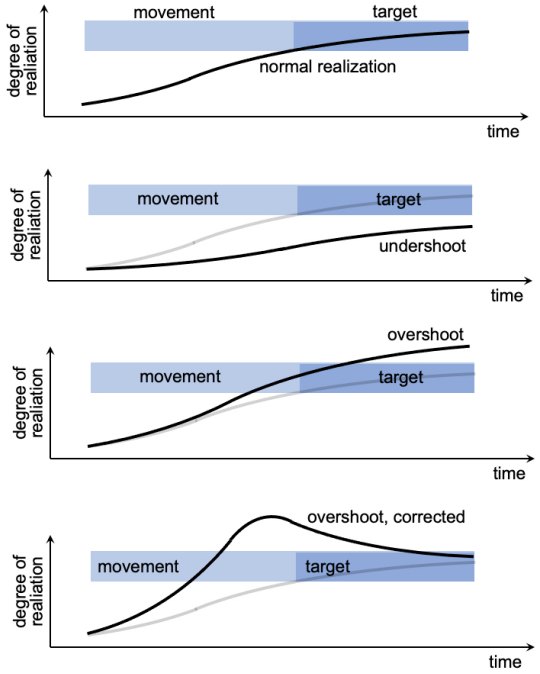

Figure 3: normal gesture, undershoot gesture, and corrected overshoot gesture

B. Simulation experiments

Five types of simulation experiment are executed: Simulation of undershoot and overshoot in case of (i) phonatory gestures (glottal and/or pulmonary gestures), (ii) vocalic gestures, (iii) consonantal gestures, and (iv) velopharyngeal gestures. Simulation of dyscoordination for (v) glottal relative to consonantal gestures.

III. RESULTS

Qualitative results for over- and undershoot of single gestures as generated by our simulation model are listed here for different types of gestures: (i) *Glottal gestures*: Undershoot and overshoot in glottal adduction (rest position of vocal folds for phonation is too wide or too narrow) was studied in the context of simple vocalic syllables like a sustained [a::]. Undershoot (rest position

is too wide) perceptually leads to breathy voice quality. Overshoot in glottal adduction (vocal folds are strongly adducted; high medial compression) leads to harsh and strained voice quality and phonation may stop. Thus, overshoot in glottal adduction gestures forces the model to overshoot the pulmonary gesture (increase subglottal pressure) in order to maintain phonation. (ii) *Vocalic gestures*: Undershoot and overshoot was studied in babbling sequences like [bababa], and [sasasa]. Undershoot results perceptually in a too central schwa-like vowel quality. Speech sounds effortless and under-articulated. In contrast, overshoot in our model leads to static and less coarticulated speech but all vowels sound clearly articulated. (iii) *Consonantal gestures*: Undershoot and overshoot was studied in the same babbling sequeunces (see above). Undershoot leads to short and imprecise productions of consonants. In few cases no closure or constriction is produced and the consonant is acoustically not present. Overshoot leads to very long constrictions or closures. Speech now sounds over-articulated. (iv) *Velopharyngeal gestures*: Undershoot and overshoot was studied in the same babbling phrases (see above). Undershoot perceptually leads to nasalized speech. Plosives and fricatives are acoustically less present, because the pressure built-up in the oral cavity is imperfect.

One experiment was conducted to study (v) *dys-coordination of consonantal and phonatory gestures* in the case of the syllable [ba]. In normal speech a phonatory gesture (glottal closing gesture) reaches its target region synchronously with the vowel (see syllable [pa] in Fig. 2: the target phase of the phonatory gesture (clos phon) starts after consonantal release of [p]). But in the case of a preceding voiced consonant (e.g., [ba]), the phonatory gesture reaches its target region earlier: normally during consonantal closure. If the glottal gesture in coordination with a pulmonary gesture now is shifted to even more earlier points in time, we get an inadequate *pre-phonation effect*, which can be transcribed as [@ba].

IV. DISCUSSION AND CONCLUSIONS

A first qualitative auditory evaluation of synthesized samples of over- and undershoot for different types of gestures as well as of temporal dyscoordination of articulatory and phonatory gestures allows an association of some of these mechanisms with types of neurogenic speech disorders. (i) Pre-phonation resulting from dysfunctions in temporal coordination of articulatory and phonatory gestures occurs in apraxia of speech. (ii) Undershoot of gestures leading to soft speech, monotonous intonation, and reduced intelligibility of speech sounds occur in hypokinetic speech. (iii) It is difficult to associate overshoot phenomena synthesized in our model with hyperkinetic speech samples. More research is needed here. (iv) It is difficult to associate under- or overshoot phenomena with ataxic dysarthria. Complex syllables often are suppressed (produced fast and slurred) in natural data while simple syllables are articulated in a normal way. That results from articulatory reorganization affecting the whole motor plan of a syllable. (v) The same applies to spastic dysarthria. If a gesture target cannot be reached in its normal time interval because movements are too slow, reorganization of the motor plan takes place and lead to an increase in duration of the movement phase of gestures and subsequently to an increase in syllable durations. (vi) In contrast, in case of flaccid dysarthria the patient does not try to reach targets because of his experience about his motor constraints (his inabilities in target reaching). The patient stays with gesture undershoot.

While this preliminary study shows the capability of our model in explaining some basic types of articulatory and phonatory settings occurring in different types of neurogenic speech disorders, a more detailed evaluation of the generated speech samples is needed for a more detailed comparison with natural speech samples.

REFERENCES

[1] A. van der Merwe, A., "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*, vol. 35, pp. 397-423, 2021.

[2] R.D. Kent, J.F. Kent, J.R. Duffy, J.E. Thomas, G. Weismer, and S. Stuntebeck, "Ataxic Dysartria," *J. Speech Lang. Hear. Res.*, vol. 43, pp. 1275-1289, 2000.

[3] F.L. Darley, A.E. Aronson, and J.R. Brown, "Differential diagnostic patterns of dysarthria," *J. Speech Hear. Res.*, vol. 12, pp. 246-269, 1969.

[4] B.J. Kröger, and P. Birkholz, "A gesture-based concept for speech movement control in articulatory speech synthesis," in *Verbal and Nonverbal Communication Behaviours* LNAI 4775, A. Esposito, M. Faundez-Zanuy, E. Keller, and M. Marinaro Eds. Berlin, Heidelberg: Springer, 2007, pp. 174-189.

[5] C.P. Browman, and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, pp. 155-180, 1992.

[6] B.J. Kröger, T. Bekolay, and C. Eliasmith, "Modeling speech production using the Neural Engineering Framework," *Proceedings of CogInfoCom 2014* Vetri sul Mare, Italy, 2014, pp. 203-208.

[7] B.J. Kröger, "On the production mechanisms of the singer's formant," *Proceedings of the 23rd International Congress on Acoustics* Aachen, Germany, 2019, pp. 4568-4575.