

*Incongruencies between phonological theory and phonetic measurement**

Doris Mücke

University of Cologne

Anne Hermes

Laboratoire de Phonétique et Phonologie, UMR 7018
(CNRS/Sorbonne Nouvelle)

Sam Tilsen

Cornell University

To assess a phonological theory, we often compare its predictions to phonetic observations. This can be complicated, however, because it requires a theoretical model that maps from phonological representations to articulatory and acoustic observations. In this study we are concerned with the question of how phonetic observations are interpreted in relation to phonological theories. Specifically, we argue that deviations of observations from theoretical predictions do not necessitate the rejection of the theoretical assumptions. We critically discuss the problem of overinterpretation of phonetic measures by using syllable coordination for different speaker groups within Articulatory Phonology. It is shown that surface variation can be explained without necessitating substantial revision of the underlying phonological theory. These results are discussed with respect to two types of interpretational errors in the literature. The first involves the proliferation of phonological categories in order to accommodate variation, and the second the rejection of a phonological theory because the model which generates its predictions is overly simplified.

1 Phonological theory and phonetic measures

1.1 Phonology as a laboratory science

A primary aim of phonology as a laboratory science is to relate language as a cognitive system to observations of the physical world. However, there is

* E-mail: DORIS.MUECKE@UNI-KOELN.DE, ANNE.HERMES@SORBONNE-NOUVELLE.FR, TILSEN@CORNELL.EDU.

This work was supported by the German Research Foundation (DFG) as part of the SFB1252 ‘Prominence in Language’ in the project A04 ‘Dynamic modelling of prosodic prominence’ at the University of Cologne. The authors thank the editors and the reviewers for their thoughtful comments and efforts towards improving our manuscript.

a problem inherent to any empirically oriented analysis paradigm. There is no clear-cut division between the abstractions of a phonological theory and the continuous variation of phonetic measures. The level of granularity often varies between analyses, and studies differ in how they interpret variation in the phonetic dimension: in some cases variation is viewed as part of the underlying phonological knowledge; in others it is merely statistical noise (Pierrehumbert *et al.* 2000). One reason for this is that phonological theories generally make use of the DISCRETE MATHEMATICS of categorical abstractions (low-dimensional representations) and relate them to the CONTINUOUS MATHEMATICS of sound patterning (high-dimensional representations; Gafos & Benus 2006). Theoretical analyses have to decide to what extent a phonetic measure can inform us about phonological structure and to what degree a phonological assumption can predict the phonetic output (Anderson 1981, Keating 1988, Ohala 1990, Blumstein 1991, Chang 2012). These decisions always depend on a theoretical model, whether explicit or implicit, of how phonological representations map to surface phonetic observations. In a phonological theory, we always have to deal with multi-faceted interactions between categorical and gradient information in order to interpret phonetic variation for theoretical purposes (Chitoran & Cohn 2009). The question arises how much deviation from structural components we want to allow for.

A different perspective comes from the theory of dynamical systems, which is able to describe relatively stable, quasi-categorical states in a completely continuous environment by using the MATHEMATICS OF NON-LINEAR DYNAMICS (Browman & Goldstein 1992, Gafos & Benus 2006, Goldstein *et al.* 2006, Tilsen 2016, Mücke *et al.* 2017, Gafos *et al.* 2020). Theories of dynamical systems are based on the assumption that the human mind steadily gravitates towards relatively stable states in a continuous space (Spivey 2007). Within a single equation, low- and high-dimensional aspects of speech are integrated by defining the (invariant) relation between (variant) parameters. Dynamical systems aim to fully integrate phonetics and phonology within a single grammatical module. Instead of assuming fixed categories and rules that derive variability from the symbolic based forms, they assume that we are dealing with attractors, the relatively stable states in a continuous space that simultaneously encode discrete and gradient aspects of speech. Even though dynamical systems in principle have the power to fully integrate phonetics and phonology, they also have to deal with the problem of determining the range of permitted speech outputs, raising the question of how much variability and stability is reasonable for a language system (Shaw *et al.* 2011).

1.2 Predefined phonetic tools

Different analytical goals may result in different conceptions of how discrete and continuous representations interact. Analytical goals therefore guide the interpretation of data from experiments, and it is commonly the case that interpretational procedures – i.e. ‘phonetic tools’ – are

highly stereotyped. The use of a predefined phonetic tool related to a specific theory is connected to the LAW OF THE INSTRUMENT, which holds that the routine usage of a familiar tool for solving different problems can limit our knowledge (if you just have a hammer, you will treat every problem as a nail; Maslow 1966). The routine usage of the same measure for different goals can lead to flawed analyses, and this is a recurring problem in different areas of experimental phonology.

There are many examples in the linguistic literature that can be discussed in light of this problem. One example is the 'rhythm class' debate, where languages are divided into distinct classes according to whether they are syllable-timed, stressed-timed or mora-timed (Abercrombie 1967, Port *et al.* 1987, Ramus *et al.* 1999, Arvaniti 2009). In this approach, local timing proportions in terms of relative consonant and vowel durations on the acoustic surface have been used to provide evidence for the different classes. However, it has been shown in further studies (Arvaniti & Rodriquez 2013, Krivokapić 2013, Tilsen & Arvaniti 2013) that many aspects of what might have been taken to be differences in rhythm are in fact related to other prosodic factors, such as speaking rate and F₀, and cannot be adequately captured by local timing patterns in many languages. This calls into question whether the rhythm-class typology is appropriate, and whether measures of segmental duration can be directly applied to the understanding of rhythmic properties across languages (Turk & Shattuck-Hufnagel 2013). It was the assumption that rhythmic structure is directly encoded in segmental durations by means of a particular linking model or interpretive tool that led to the overinterpretation of such measures.

Another example involves tonal alignment, where the temporal coordination of pitch movements with consonants and vowels of the segmental string is investigated. In the autosegmental-metrical approach, tones are associated with tone-bearing units in the segmental string, such as stressed syllables in German pitch accents. Tonal alignment research extends this concept by developing the segmental anchoring hypothesis, which involves the measuring of patterns of the co-occurrence of pitch movements with boundaries of the segmental string in the acoustic dimension (Arvaniti *et al.* 1998, Ladd *et al.* 1999, Ladd *et al.* 2000, D'Imperio *et al.* 2007, Prieto & Torreira 2007, Mücke *et al.* 2009). Tonal ASSOCIATION is categorical and low-dimensional, while tonal ALIGNMENT is continuous and high-dimensional. Originally, segmental anchoring and the related measures of co-occurring events in the tonal and segmental string were not intended to be direct reflexes of phonological categorisation. However, the measures applied (capturing latencies between F₀ turning points in rising and falling pitch accents and segmental boundaries of consonants and vowels in the accented syllable) have led to controversial implementations in phonology. For example, Prieto *et al.* (2005) suggest accounting for the variation found in Romance languages by augmenting the concept of tonal alignment with secondary associations. The idea is to have a primary association between tone and tone-bearing unit, and in

certain cases to have a secondary association between tones and edges of prosodic constituents. The secondary associations push the realisation of the tone movement towards the edge of the prosodic category. For this kind of phonological implementation, however, Ladd (2006, 2008) points out that a theory which posits fine-grained categories of phonetic alignment patterns runs the risk of proliferating phonological categories.

1.3 Interpretational error types in the analysis of syllable coordination

The example of misinterpretation of a phonetic measure that we focus on in this paper involves the syllable-coordination paradigm assumed by Articulatory Phonology. In this research paradigm, a phonetic effect known as the C-centre is used as a key diagnostic for different forms of phonological syllable organisation. However, with every new C-centre study, new surface patterns are identified which do not appear to conform to theoretical predictions.

Within Articulatory Phonology, it is assumed that distinct phonological syllable parses such as simple (non-branching) and complex (branching) onsets correspond to different organisations of consonants and vowels in the articulatory domain (e.g. Browman & Goldstein 2000, Shaw *et al.* 2009, 2011, Gafos *et al.* 2010, Marin & Pouplier 2010, Hermes *et al.* 2013, Hermes *et al.* 2017).

Depending on their position in the syllable, consonantal and vocalic gestures are coordinated differently with respect to one another, and it is hypothesised that the underlying phonological organisation varies with syllable complexity, for example between CV and CCV. Furthermore, Articulatory Phonology claims that there are two distinct phonological forms of organisation for word-initial consonant clusters: (i) complex organisation, in which both consonants are associated with the same syllable, and (ii) simplex organisation, in which the initial consonant is extra-syllabic, i.e. less closely associated with the syllable projected by the following vowel. Hermes *et al.* (2013) provide evidence from Italian consonant clusters, which show a complex organisation for obstruent–liquid clusters (e.g. /pr/ in *prima* ‘first’) and a simplex organisation for sibilant–obstruent clusters (e.g. /sp/ in *spina* ‘thorn’).

Empirically, it has been observed that when a consonant is added to the beginning of a word to form a complex onset, the prevocalic consonant is shifted towards the vowel to make room for the added consonant. This is the empirical pattern referred to as the C-centre effect, and has been taken to provide phonetic evidence for complex organisation in phonological theory. The C-centre effect has been reported for clusters like /pl/ and /kl/ in Polish (Mücke *et al.* 2010, Hermes *et al.* 2017), American English (Browman & Goldstein 1988, Honorof & Browman 1995, Marin & Pouplier 2010, Walzl & Marin 2010), Italian (Hermes *et al.* 2013), French (Kühnert *et al.* 2006) and Romanian (Marin & Pouplier 2014). The C-centre effect is a phonetic reflex of an underlying phonological

syllable parse, and is usually diagnosed by measures of articulatory overlap between initial consonants and the following vowel (C-centre measures). Due to compression, the overlap between the vowel and the prevocalic consonant is greater in complex onsets than in simplex onsets. However, variability due to prosodic and segmental factors can affect the overlap patterns (Goldstein *et al.* 2006, Goldstein *et al.* 2009, Shaw *et al.* 2011, Pastätter & Pouplier 2015, Hermes *et al.* 2017), and can even block the increase in overlap between C₂ and V in a C₁C₂ sequence. This can lead to deviations from the C-centre timing pattern in languages with complex onsets, which could be misinterpreted as evidence for simplex organisation. This is especially the case for stop–lateral sequences in languages such as German, which are claimed to be organised as branching onsets in the relevant phonological literature, but do not show the expected C-centre effect in the phonetic output. Pouplier (2012) and Brunner *et al.* (2014) show that the German cluster /pl/ fails to reveal a C-centre timing pattern, despite the fact that phonological considerations indicate that /pl/ is a complex onset in German (Wiese 1996). The same is true for /pl/ in Hebrew and Montréal French, where Tilsen *et al.* (2012) also found no C-centre coordination for /pl/ in the kinematic signal, even though both languages are expected to allow complex onsets (Bolozky 1997 for Hebrew; Kühnert *et al.* 2006 for French). When phonological theory and phonetic measures are incongruent, we can find interpretational errors in the literature. Brunner *et al.* (2014), for example, claim for their German findings that the observed timing measures are influenced by segmental composition and coarticulation in the phonetic output rather than reflecting the underlying phonological structure. This implies that the C-centre approach should be rejected as a reliable measure of phonological theory in general (or in a specific language or for a specific sequence of consonants). Another possibility is to assume on the basis of the applied measures that Hebrew does not have complex onsets phonologically (Tilsen *et al.* 2012). In this case the phonological classification for this specific language should be changed from complex to simplex syllable parses. However, both interpretations can be understood as typical cases of the interpretational error types described above.

The uncritical use of a measurement to draw theoretical conclusions can lead to phonological misinterpretations. This is a recurring problem, which emerges particularly within established research paradigms, rather than in new phonological theories. The routinised use of a research tool (i.e. drawing inferences about syllable organisation from the presence or absence of the C-centre effect) can lead to apparent incongruencies between theoretical expectations and empirical observations. These discrepancies are usually tolerated until the ‘elephant in the room’ becomes too big to be ignored. Even though we might be aware of this problem, there still seems to be uncertainty of how to deal with it. There are likely numerous cases where we could have asked ourselves whether problems have arisen from the use of an inadequate instrument (‘This measure has been frequently used; why not for this goal?’), or

whether we were using overly simplistic models of how surface variation is generated by our phonological theories ('What happened to my phonological form?'). The two types of interpretational errors are summarised in (1).

(1) a. *Revision of the theory*

Phonological representations are adjusted or phonological categories are added, in order to account for surface variation.

b. *Rejection of the theory*

Incongruencies between empirical observations and theoretical predictions are taken as evidence that the theory should be rejected.

Crucially, we argue that, for both types of interpretational error, the cause of the problem is in many cases an overly simplistic or incorrect model of how the theory generates empirical predictions.

1.4 Aim of the present study

In this paper we discuss the problem of incongruencies between phonological theory and phonetic measures in experimental approaches. Our overarching goal is to argue for more caution and analytical rigour in the assessment of phonological theories via phonetic measurements, by paying closer attention to how they are linked. In doing so, we will focus on deviations of surface timing patterns from theoretical predictions. More specifically, we adapt examples taken from the C-centre approach, framed in the theory of non-linear coupled oscillators, which forms part of Articulatory Phonology.

The empirical data we examine here were obtained from stop–lateral sequences in German, recorded with a 3D electromagnetic articulograph. We chose stop–lateral patterns since they are described in the literature as being problematic, violating the predictions of articulatory overlap predicted by complex onset organisation (Poupier 2012, Brunner *et al.* 2014). We compared variation in surface patterns of different German populations that have been described to show changes in the speech motor control system: younger *vs.* older speakers (Hermes *et al.* 2018) and pathological speech from Essential Tremor patients treated with deep brain stimulation with age-matched healthy control speakers (Mücke *et al.* 2018, Hermes *et al.* 2019). The results will be discussed within Articulatory Phonology, but also with respect to general problems of incongruencies between theoretical predictions and surface patterns.

To assess whether the theory of Articulatory Phonology can account for variation in the data, we use a computational model that employs generalised COUPLING STRUCTURES to generate timing relations in syllable onsets. We tested whether the parameters of this model can be optimised to account for surface variations. In pursuing this, we distinguish between categorical parameters such as coupling structure, which determines network topologies for simplex and complex onset syllable parses, and

two different gradient parameters, which include coupling strength and corrections for biomechanical interactions of articulators. We note that the gradient coupling-strength parameters are implicit in simpler models, but are commonly treated as fixed. In this study, a novel parameter for biomechanical correction is introduced in the models, which are evaluated according to their ability to fit the empirical data. We expect to find the following:

(i) There are systematic incongruencies between phonological predictions and surface patterns in stop–lateral sequences such as /pl/ and /kl/ in German when applying the Articulatory Phonology model in its most basic form. In kinematic measures, the basic Articulatory Phonology model predicts an increase in overlap between the prevocalic C and the following V when a C is added to the beginning of the word, but this is not observed in empirical studies (e.g. for stop–lateral sequences in German; Pouplier 2012, Brunner *et al.* 2014).

(ii) The differences between predictions and observations will become even larger when coordination patterns of different populations are included (ageing and pathological speech).

(iii) If the Articulatory Phonology model is extended to allow for asymmetries in consonant–vowel COUPLING STRENGTH, the congruency between the phonological prediction of a complex onset parse and the phonetic output pattern will be considerably improved. We view this coupling-strength parameter as part of the grammatical knowledge of the speaker.

(iv) If the Articulatory Phonology model is further extended to account for biomechanical interactions between articulators, the congruency between theoretical prediction and output pattern will again be improved. This parameter is motivated on the basis of physical interactions between the tongue, jaw and lips.

(v) Allowing for dynamic adjustments of model parameters within the same phonological structure leads to better results than merely changing the categorical ones that relate to network topology.

We note here that our immediate aims in pursuing the above analyses are neither to argue for a particular phonological theory nor to argue for a particular linking model. Rather, our principal aim is to demonstrate how interpretation of empirical data necessitates critical examination of the model that links a theory to its predictions.

2 Coupled oscillators: theory, model and empirical assessment

2.1 The coupled oscillators theory of Articulatory Phonology

Articulatory Phonology is a theory that decomposes speech into a set of potentially overlapping units, ARTICULATORY GESTURES (Browman & Goldstein 1986, 1992, 2000). Articulatory gestures define coordinated

articulatory actions which achieve a linguistic goal such as the full closure of the tongue tip at the alveolar ridge to produce the oral closure for the speech sound /t/. Since gestures overlap in time, they encode a great amount of context-dependent variability, reflecting functional synergies of the articulators moving towards different competing attractors (Fowler *et al.* 1980, Saltzman & Kelso 1987, Saltzman & Munhall 1989, Browman & Goldstein 1992, Hawkins 1992). In the intervocalic consonant /t/, the goal for the tongue-tip closure at the alveolar ridge is invariant, but the way the tongue tip travels in the physical representation differs in utterances such as /ata/ and /iti/, due to the different starting conditions of low and high vowels.

As a dynamic theory, Articulatory Phonology fully integrates low-dimensional descriptions (the gesture as a discrete phonological unit) and high-dimensional descriptions (the gesture as a continuous physical articulatory action) in a unified system, by using laws to describe the speech system's behaviour in terms of differential equations. While the laws for modelling a specific utterance are invariant, the physical output is not (see Browman & Goldstein 1992, Kelso 1995, Gafos & Benus 2006, Spivey 2007, Gafos *et al.* 2014, Mücke *et al.* 2017). Changing the value of a gesture's parameter set changes the temporal and/or spatial properties of the physical articulatory action, and therefore the outcome measurable on the surface. In a dynamical system there is, strictly speaking, no 'mapping' between phonological and phonetic information, i.e. we do not have modules for discrete phonological and continuous phonetic information of grammatical knowledge (Ohala 1990). Moreover, dynamical systems integrate these aspects by the use of non-linear mathematical equations. The equations define the relations between parameters, and these relations represent the invariant part of speakers' knowledge, while the concrete parameter values generate gradience in the output. Dynamic systems are not based on what we describe as categories in the traditional sense. Instead, they use attractors, which operate in a completely continuous environment, rather than fixed categories. The attractors evolve over time towards quasi-categorical states. For example, a quasi-categorical state can be a coupling structure for a complex or a simplex onset parse that coordinates the phasing between articulatory gestures in a coupled oscillator network (Browman & Goldstein 2000, Cho 2006, Nam *et al.* 2009, Marin & Pouplier 2010, Shaw *et al.* 2011, Hermes *et al.* 2013, Gafos *et al.* 2014, Hermes *et al.* 2017).

Within this network, each gesture is associated with an oscillator (or clock), and the oscillators are coupled to one another in a pairwise fashion. Coupling structures (i.e. in-phase and anti-phase modes and the respective coupling forces) are phonological in nature, while the output of a coupling network has phonetic consequences: the coupling structure determines the initiation and coordination of articulatory speech gestures, and can lead to the C-centre effect in the phonetic output of complex onsets. In the planning process of an utterance, coupling between the oscillators forces them to settle into a stable timing pattern. In execution, the

oscillators then function to trigger the initiation of a specific gesture that they are coupled to. The model is highly constrained, in that there are only two available coupling modes: in-phase (relative phase 0° of a gesture's oscillator) and anti-phase (relative phase 180° of a gesture's oscillator).

Figure 1 schematises the internal syllable coordination of consonants and vowels in CV, VC, CCV and C.CV syllables. The top of the figure shows structural representations of syllables in autosegmental phonology (Hyman 1975). In the middle, syllables are represented as network structures representing relations between coupled oscillators. The solid lines correspond to the in-phase mode (articulatory movements are initiated at the same time), and dashed lines correspond to the anti-phase mode (movements are initiated sequentially). At the bottom of the figure, gestural scores display gestural activation intervals, which are the phonetic outputs of the corresponding phonological coupling structures. Each box in a gestural score defines the gestural activation interval from the initiation of the movement to the achievement of the target. When a movement is initiated, the corresponding articulator starts to move towards a target, such as a full closure at the alveolar ridge for the production of /t/, and the end of a box indicates that the gesture is deactivated. The gestural scores show that consonantal and vocalic movements overlap in time, encoding coarticulation. The patterns of organisation of the syllable types in Fig. 1 can be characterised as in (2).

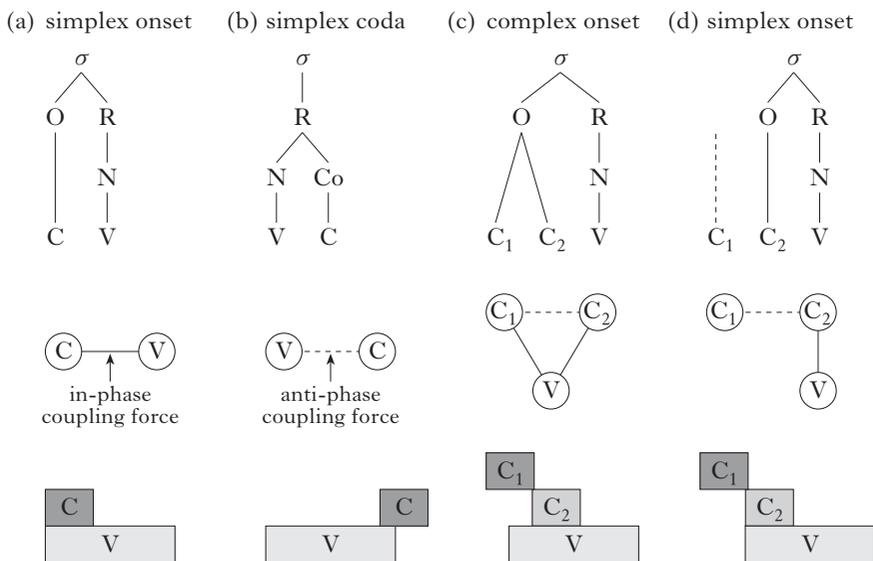


Figure 1

The organisation of CV, VC, CCV and C.CV syllables.

The figure shows autosegmental tree structures (top), coupling graphs (middle) and gestural scores (bottom).

- (2) a. CV syllables are associated with in-phase coupling. The consonantal and vocalic gestures start at approximately the same time, but the vocalic movement is executed more slowly. This results in a consonant–vowel sequence on the acoustic surface.
- b. In a VC syllable, the gestures are associated with anti-phase coupling, resulting in sequential activation, i.e. staggered initiation of timing.
- c. In a CCV syllable with a branching onset, the two consonants are in-phase with the vowel, but anti-phase with each other. This leads to a leftward shift of the initial C away from the following V, and a rightward shift of the prevocalic C towards the V. This pattern is referred to as the C-centre effect. It is usually assumed that the C-centre effect exhibits symmetrical shift patterns of C_1 and C_2 , implying balanced coupling forces. The occurrence of this pattern is language-specific.
- d. An alternative organisation is a C.CV structure, with a non-branching onset and extraprosodic initial consonant. Only the immediately prevocalic consonantal gesture is in-phase coupled to the vocalic gesture in this case. This leads to a timing pattern in which the initiation of the immediately prevocalic gesture is synchronised with the initiation of the vocalic gesture, just as in a simple CV syllable.

Coupling topologies in the coupled oscillators model are categorically different forms of phonological organisation of articulatory gestures (Browman & Goldstein 2000). Languages differ in the coupling topologies they use for syllable affiliation (simplex and complex onset parse), and such topologies have to be learned. The traditional branching onset structure, as found in German, English and Polish, for example, corresponds to a topology in which both consonantal planning oscillators are coupled to the vowel (Fig. 1c), whereas a non-branching structure, as in Tashlhiyt Berber and Moroccan Arabic, corresponds to only the immediately prevocalic consonant being coupled to the vowel (Fig. 1d; Shaw *et al.* 2011, Pouplier 2012, Hermes *et al.* 2017).

The coupled oscillators model is relatively low-dimensional: there are just two types of coupling forces, with several different network topologies. ‘Network topology’ refers to the pattern of coupling between gestural planning oscillators. In a prototypical form, it is assumed that the coupling forces are equal in strength, and this produces a C-centre pattern with symmetrical shifts of C_1 away from V, and C_2 towards V. Figure 2 provides examples of consonantal gesture timing for the cluster /pl/ in Polish and German, taken from Mücke *et al.* (2010). Both languages are assumed to have complex onset organisation. In Polish, speakers indeed produce the expected prototypical symmetrical pattern, fully consistent with the theoretical model. C_1 is shifted away from the V to decrease the overlap, and C_2 is shifted towards V to increase the overlap. However, German speakers do not show a symmetrical shift pattern. There is a considerably smaller rightward shift of C_2 , and the shift of C_2 is considerably

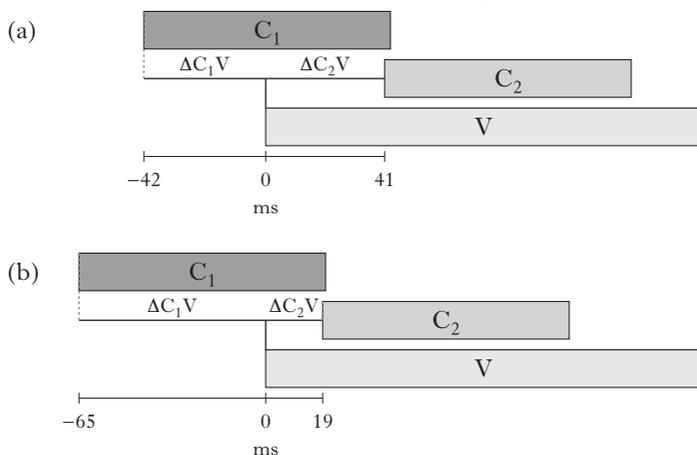


Figure 2

C-centre organisation of the cluster /pl/: (a) prototypical C-centre effect (e.g. Polish); (b) ambiguous C-centre effect (e.g. German).

smaller than the shift of C_1 . Even though the surface patterns in Polish and German are assumed to be derived from the same phonological complex onset parse, the timing of consonantal and vocalic movements in the two languages differs considerably.

This kind of variation is not captured in standard implementations of the coupled oscillators model. Moreover, the difference between complex and simplex onset cluster topologies cannot generate the contrast between Figs 2a and b. This results in an interpretational dilemma. On the one hand, we might conclude that the German pattern reflects simplex onset organisation, because it does not exhibit a prototypical C-centre effect. Alternatively, we might reject the coupled oscillators model entirely, since its predictions are not consistent with our expectations that German /pl/ cluster is a complex onset. We think that both of these interpretations are misguided. To show why, we develop an extended computational model that accounts for variation in the German timing pattern, while preserving the network topology used for complex organisation. We demonstrate that the absence of a prototypical symmetrical C-centre pattern ('the magic moment measure'; Vatikiotis-Bateson *et al.* 2014) does not necessitate rejection of the underlying complex onset coordination.

2.2 Implementation of the theory: modelling surface timing patterns

In this section we describe a computational implementation of the standard coupled oscillators model, and then introduce two extensions of the model. These extensions allow the model to better fit empirical data, without necessitating a rejection of the hypothesis of complex

organisation. Our aim here is to provide the reader with sufficient background to understand the standard model and our extensions; we employ a number of visualisations to accomplish this. More mathematical presentation and ancillary detail are given in the Appendix.¹

2.2.1 The standard coupled oscillators model: balanced coupling. The model implementation we use is based on a standard version of the coupled oscillators model of Articulatory Phonology (Saltzman *et al.* 2008, Tilsen 2017). Figs 3 and 4 depict the model in the production of a complex onset CCV syllable. In these figures, C_1 , C_2 and V do not refer to segments, but are labels for consonantal and vocalic articulatory gestures and gestural planning oscillators. Figure 3 shows how the states of planning oscillators generate a pattern of relative timing of gestural activation. For simplicity, we omit the glottal and velic gestures, and assume that C_1 , C_2 and V gestures specify targets for oral tract variables such as tongue tip,

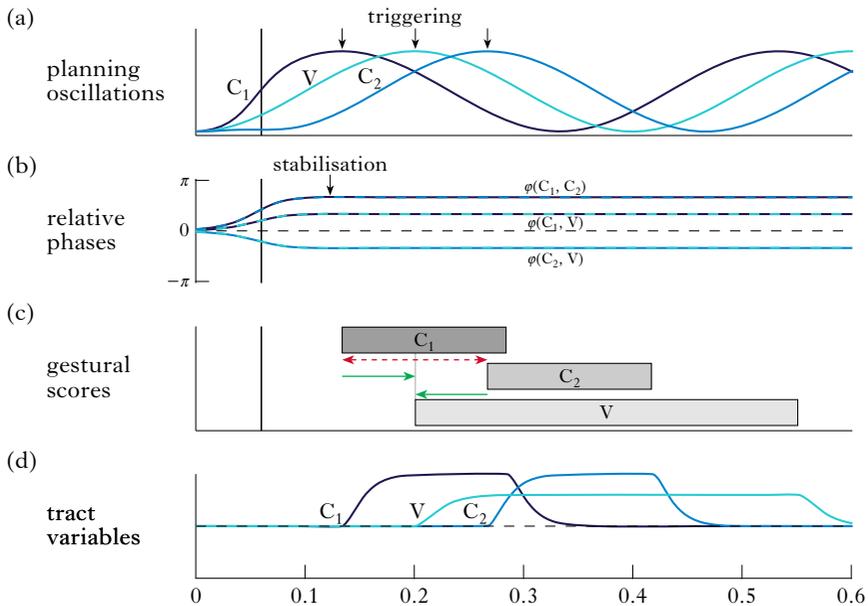


Figure 3

Overview of the coupled oscillators model of the C-centre effect: (a) planning oscillations over time (arrows indicate when each oscillator triggers activation of the corresponding gesture); (b) relative phases over time (after stabilisation the oscillators trigger the initiation of gestural activation); (c) gestural scores; (d) tract variables.

¹ The appendix is available as online supplementary materials at <https://doi.org/10.1017/S0952675720000068>.

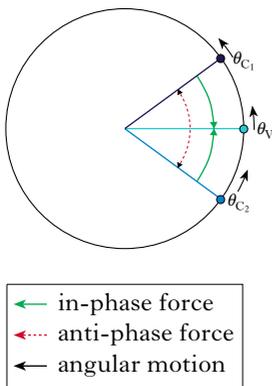
tongue body and lip aperture. In the coupled oscillators model, as explained in §2.1, each gesture is associated with a gestural planning oscillator. It is important to understand that gestures and planning oscillators are different types of systems: gestures are systems which transition between active and inactive states, and influence the target state of the vocal tract; planning oscillators are systems which intrinsically oscillate, and determine when gestures become active.

The waves associated with planning oscillators in the production of a CCV syllable are shown in Fig. 3a. These are labelled C_1 , V and C_2 respectively. When each wave first reaches its peak, the activation of the corresponding gesture is triggered. For a complex onset CCV syllable, C_1 is triggered before V, which is in turn triggered before C_2 . This temporal ordering is evident from a comparison of the triggering arrows in Fig. 3a and the onsets of the gestural activation intervals in Fig. 3c: the arrows in (a) indicate the point in time corresponding to the start of gestural activation intervals (the boxes in (c)). A precondition for triggering is the stabilisation of the relative phases of the planning oscillators, shown in Fig. 3b. Stabilisation is achieved by relative phase coupling forces, which we examine in Fig. 4. For the moment, the reader should simply understand that the relative phases shown in Fig. 3b ($\varphi_{C_1C_2}$, φ_{C_1V} , φ_{C_2V}) are determined by the structure of coupling relations between the planning oscillators. Recall that different structures ('topologies') were shown in Fig. 1, for CV, VC, complex onset CCV and simplex onset C.CV forms. Here we examine a model which employs the complex CCV topology. We consider a network of different clocks (or oscillators) that determine the start of several vocal tract movements relative to each other. It is important to realise that the relative TIMING of the initiation of gestural activation is directly related to the RELATIVE PHASES of the planning oscillators (cf. Tilsen 2018).

Consider also that the timing pattern shown in Fig. 3c is SYMMETRIC: the initiations of the C_1 and C_2 gestures are equally displaced from the initiation of the V gesture in opposite directions in time. The underlying cause of this symmetry is a balance between coupling forces, which we discuss below. We will refer to this symmetrical variety of coordination as BALANCED COUPLING. To emphasise the point that the timing of gestural initiation is determined by relative phases of planning oscillators, which are in turn determined by coupling forces, we illustrate the temporal effects of these forces in Fig. 3c. The solid arrows show that the in-phase coupling forces act to bring the initiations of C_1 and C_2 closer to the initiation of V; the broken arrow shows that the anti-phase coupling force acts to make the initiations of C_1 and C_2 more distant in time. The reader should note that the coupling forces act on planning oscillators, *not* on the gestures themselves; the timing of gestural activation is indirectly determined by the oscillators, because gestures are activated (triggered) when oscillators reach a particular phase.

Finally, Fig. 3d shows that empirically observable state variables of the vocal tract, i.e. TRACT VARIABLES, are driven toward gesture-specific targets when gestures are activated. In the absence of activation, the tract variables

(a) phase circle



(b) relative phase potentials and forces

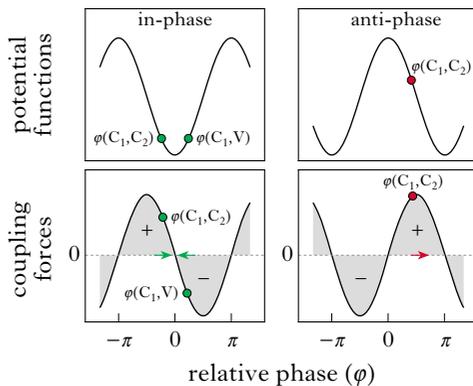


Figure 4

Coupling forces in the coupled oscillators models of the C-centre effect:

- (a) planning oscillator phases on the unit circle and the influence of coupling forces ($\varphi(C_1, V) = \theta_{C_1} - \theta_V$, $\varphi(C_2, V) = \theta_{C_2} - \theta_V$, $\varphi(C_1, C_2) = \theta_{C_1} - \theta_{C_2}$;
 (b) in-phase and anti-phase potential functions and coupling forces.

return to neutral values which are similar to the configuration of the vocal tract during the production of a neutral vowel like schwa. Specific tract variables and values are not indicated in the figure. For a concrete example, the reader can imagine that C_1 is a labial closure gesture, in which case C_1 specifies a target value of the lip aperture tract variable, where the target corresponds to the lips being closed. When C_1 becomes active, the tract variable is driven toward this target value. When C_1 deactivates, the tract variable returns to a neutral value (Saltzman & Munhall 1989).

Planning oscillators, unlike gestures, are systems which exhibit an intrinsic oscillation. The state of a planning oscillator can be readily visualised as the angle of a point moving around a unit circle, as in Fig. 4a. In technical contexts, the state of an oscillatory system is often called a PHASE ANGLE (θ), and radians are used rather than degrees. As a matter of convenience, phase angle is referred to simply as PHASE, or θ . Phase is periodic on the interval $[0, 2\pi]$. As implied by the arrows outside the circle in Fig. 4a, the phases of the C_1 , V and C_2 planning oscillators (θ_{C_1} , θ_V , θ_{C_2}) revolve around the unit circle. It is important to recognise that the oscillators always revolve around the circle in this manner.

In addition to the ever-present revolution of phase, coupling forces can exert effects on oscillator phases. These effects are often manifested as a transient slowing down or speeding up of the revolution (i.e. changes in angular velocity). If only in-phase coupling were present in this example, the systems would evolve to have exactly the same phase, and the gestures would be activated at the same time. If only anti-phase coupling were present, the systems would evolve to be maximally separated around the unit circle (i.e. separated by a distance of $\pi/3$ radians, or

120°), and the corresponding gestures would be activated sequentially. But for a complex CCV syllable, both in-phase and anti-phase forces are hypothesised to be present, and this can lead to a pattern of system phases such as the one shown in Fig. 4b. This leads to a symmetrical shift of C_1 and C_2 towards and away from the V (the corresponding gestures are activated in the order C_1 -V- C_2).

The coupled oscillators model does not merely stipulate a stable relative phase pattern. Instead, the relative phase pattern emerges as a consequence of relative phase coupling forces, under fairly general assumptions (see the Appendix). Figure 4b shows the sinusoidal potential functions and associated forces, for both in-phase and anti-phase coupling. There are several key points to make regarding these functions. First, the horizontal axis in all cases is RELATIVE PHASE (φ), i.e. the difference between phase angles. Second, the forces in question are forces *on* relative phase, i.e. the forces act to increase or decrease φ . These actions on φ are translated to changes in phase velocity (see the Appendix). Third, the force functions are the negative derivatives of the potentials, with respect to relative phase. Hence, when a potential function decreases as relative phase increases, the force is positive. When a potential function increases with relative phase, the force is negative. These areas are shaded and labelled '+' and '-' in Fig. 4b. Furthermore, at the minimum of a potential function, the force is zero. Consequently, the change of φ over time can often be predicted from the potential function by imagining the relative phase to be a marble rolling in a bowl with a sticky surface. In this metaphor, the bowl defines all possible values of a continuous phase space. After the system is set into motion, the marble rolls downwards in the bowl. The bottom of the centre of the bowl is comparable to the attractor of the dynamic system defining a linguistic target (i.e. the equilibrium position), as in-phase and anti-phase modes. The fourth key point is that a stable equilibrium is a positive-to-negative zero-crossing in the force function. This is reflected by the fact that the minimum of the in-phase potential is $\varphi = 0$, while the minimum of the anti-phase potential is $\pm\pi$ radians. Lastly, in the absence of other forces, the relative phase will always move to a local minimum in the potential function. If there is a competition between several target attractors (i.e. the marble is in several 'bowls' at the same time), the position where the marble comes to rest may not be the bottom of any particular bowl. Indeed, this is the case in the C-centre effect: in-phase forces between the vowel and each consonant are opposed by an anti-phase force between consonants.

2.2.2 Model extensions: imbalanced coupling and biomechanical correction. In this section we show how the model can be generalised to better fit empirically observed deviations from the prototypical symmetrical shift pattern in CCV syllables. To accomplish this, two extensions are introduced to the model: (i) imbalance of coupling strength, and (ii)

coarticulatory effects due to biomechanical shortening. We show that incorporating these two extensions leads to better empirical coverage.

Regarding (i), coupling strength imbalance, notice that the standard model does have coupling strength parameters, but these are artificially constrained, such that C_1 and C_2 are coupled in-phase to V with equal strength. Hence generalising the model to allow for unequal (or imbalanced) coupling does not require the introduction of a new parameter *per se*; it merely makes use of existing parameters by relaxing a constraint on those parameters. The benefit of this is that it allows the model to fit ‘ambiguous C-centre patterns’ such as the one shown in Fig. 2b. Note that gradient adjustment of a free parameter is not comparable to the introduction of a new structural component or a change in network topology.

The second extension – coarticulatory effects due to biomechanical shortening – is also readily justifiable on the basis of known interactions between the vocal organs of the jaw, tongue and lips. In discussing these below, we note that our empirical data provide indirect support for the existence of such effects.

In order to evaluate the performance of our extended model, we compare it to a standard simple model and a complex model (with balanced coupling). The comparison is based on the ability of the models to generate empirically observed timing patterns in a word-initial CCV form. We also consider a structurally heterogeneous model which allows for either simple or complex organisation on a by-speaker by-condition basis, and a heterogeneous constrained model which derives from the hypothesis that different speaker populations (i.e. younger *vs.* older speakers, or patients *vs.* controls) uniformly employ either simple or complex balanced organisation. A total of ten models are compared; these are summarised in Table I.

network topology	coupling-strength parameters	biomechanical correction
simplex	<i>n/a</i>	no
simplex	<i>n/a</i>	yes
complex	balanced	no
complex	balanced	yes
complex	imbalanced	no
complex	imbalanced	yes
heterogeneous unconstrained	<i>n/a</i>	no
heterogeneous unconstrained	<i>n/a</i>	yes
heterogeneous constrained	<i>n/a</i>	no
heterogeneous constrained	<i>n/a</i>	yes

Table I

Summary of models.

The simplex models have the network topology in Fig. 1d above. This coupling structure involves an in-phase relation between C₂ and V, but C₁ is not directly coupled to V. The coupling structure allows for simplex C.CV patterns only. This is accomplished by setting the C₁-V coupling strength parameter to 0, as shown in (3). The complex balanced models presented in Figs 3 and 4 have the network topology shown in Fig. 1c. To allow for an imbalance in coupling strength, the strengths of C₁-V and C₂-V in-phase coupling forces can differ in the complex imbalanced models. If the C₁-V coupling strength is less than the C₂-V coupling strength, there is no longer a symmetrical shift pattern in the phonetic output. Even though the underlying topology of coupling relations is the same as in the complex balanced model, the complex imbalanced model allows for the shift of C₁ away from V to be greater than the shift of C₂ towards V.

To further illustrate the differences between the models, we discuss the parameterisation in more detail here, in relation to the matrices in (3).

(3) *Coupling-strength parameter matrices*

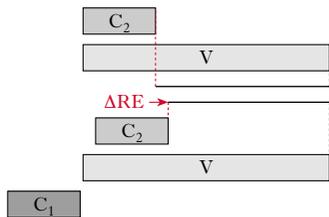
	a. <i>simple coordination</i>			b. <i>complex balanced</i>			c. <i>complex imbalanced</i>				
	C ₁	C ₂	V	C ₁	C ₂	V	C ₁	C ₂	V		
C ₁		<i>b</i>	0	C ₁		<i>b</i>	<i>a</i>	C ₁		<i>b</i>	<i>a</i> ₁
C ₂	<i>b</i>		<i>a</i>	C ₂	<i>b</i>		<i>a</i>	C ₂	<i>b</i>		<i>a</i> ₂
V	0	<i>a</i>		V	<i>a</i>	<i>a</i>		V	<i>a</i> ₁	<i>a</i> ₂	

In standard implementations of the coupled oscillators model of the C-centre effect, coupling forces are assumed to be balanced in two ways: (i) the strength of the anti-phase force (*b*) and the average of the in-phase forces (*â*) are equal: (*b/â* = 1), where *â* = (*a*₁ + *a*₂)/2, and (ii) both consonantal gestures are coupled in-phase to the vocalic gesture with equal strength (*a*₁ = *a*₂). The coupling-strength parameters of the standard model and the alternatives we explore are represented in (3). We refer to the anti-phase to in-phase ratio (*b/â*) as the STRENGTH OF ANTI-PHASE COUPLING RELATIVE TO IN-PHASE COUPLING, and the difference between *a*₁ and *a*₂ (i.e. *a*₁ - *a*₂) as the COUPLING IMBALANCE.

The simple coordination model in (3a) lacks coupling between C₁ and V, corresponding to the topology in Fig. 2d. In this model there is no interaction between the forces which determine C₁-C₂ phasing and C₂-V phasing. Indeed, the values of these parameters only influence how quickly the model will stabilise; the stabilised pattern is always one in which $\varphi_{C_1C_2} = \pi$ and $\varphi_{C_2V} = 0$. The model can fit variation in timing of C₁ and V by allowing the oscillator frequencies to vary (see §2.1.1 and the Appendix), but it will always generate synchronous activation of C₂ and V, i.e. a 0 ms difference.

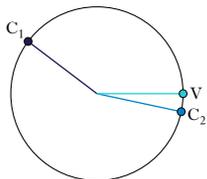
The complex balanced model in (3b) is constrained by the condition that the coupling strengths of C₁ and C₂ to V are equal. This is represented in the matrix as the presence of a single parameter for in-phase coupling

(a) imbalanced coupling



	C ₁	C ₂	V
C ₁		-	+
C ₂	-		+
V	+	+	

imbalance of in-phase coupling:
C₂V > C₁V



(b) biomechanical shortening

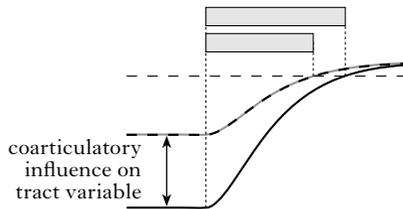
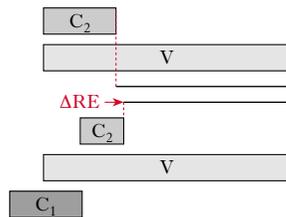


Figure 5

Extensions of the coupled oscillators model which can account for asymmetric shifts. (a) Imbalance of in-phase coupling strengths (C₂V > C₁V) results in a smaller ΔRE than in the balanced coupling model. (b) Biomechanical interaction from coarticulation of C₁ and C₂ results in a ΔRE which underestimates the shift of C₂ gestural initiation.

strength, *a*. Under the assumption of balanced coupling and equally strong in-phase and anti-phase coupling, the stabilised relative phases of C₁ and C₂ to V are $\varphi = \pm\pi/3$ (see Tilsen 2017 for a derivation of this). The complex balanced coupling model always generates a symmetric C-centre effect. By ‘symmetric’ we mean that the leftward shift of C₁ relative to V is equal to the rightward shift of C₂ relative to V. We refer to these as LE and RE shifts (or ΔLE and ΔRE) respectively, since C₁ moves toward the left edge of the word, and C₂ moves toward the right edge.

The complex imbalanced model in (3c) allows for C₁ and C₂ planning oscillators to be in-phase coupled to V with different strengths, *a*₁ and *a*₂ respectively. This lets the model generate asymmetric C-centre patterns, as in Fig. 2b. Another example is shown in Fig. 5a, where the C₁-V coupling strength is weak relative to the C₂-V coupling strength. This results in greater temporal proximity between initiation of C₂ and V than between initiation of C₁ and V. Imbalanced coupling can therefore fit departures from symmetric shifts. Note that none of the models can generate patterns in which a RE shift appears to result in initiation of C₂ before V, but to some extent our other mechanism – biomechanical correction – can account for such patterns (see discussion below).

To allow for the possibility that speakers may differ in whether they adopt a simple or complex organisation for different tasks and/or combinations of gestures, we explored two model variants which allowed for heterogeneous topologies. In the heterogeneous unconstrained model, the best fitting simple or complex balanced model was selected for each speaker/condition/target in our datasets. This amounts to allowing different speakers to adopt different coupling topologies in different conditions or for different targets.

We also examined a heterogeneous constrained model in which the following structural restrictions were imposed for the different speaker groups analysed in the present study. We investigate syllable organisation patterns for older and younger speakers (the ageing group) and for pathological speech comparing healthy controls and Essential Tremor patients treated with deep brain stimulation (the DBS group). For the ageing dataset, older speakers were assumed to use a complex balanced organisation, and younger speakers a simple organisation. For the DBS dataset, the assumption was that the control group used a complex balanced organisation, and the patient group a simple organisation. Hence, in the heterogeneous constrained models, the organisation of control was always the same for both targets (/kl/ and /pl/) and all speakers within a subject population.

To model coarticulatory effects due to BIOMECHANICAL SHORTENING, we incorporated an additional parameter which adjusts the ΔRE generated by the model. The adjustment was constrained to be from 0 to 40 ms. This parameter and its constraint can be justified as follows. There is always an interaction of lingual and labial consonant articulator movements, due to their shared connection with the jaw. For our datasets we used triples of target words in German such as *Klima*, *Lima*, *Kima* and *Plina*, *Lima*, *Pina*, as discussed in detail in §3.1.1 below. The distance that the tongue tip moves for the alveolar lateral /l/ is shorter in *Klima* and *Plina* than in *Lima*, even though the underlying goal for /l/ is invariant. The reason for this is that the jaw is already higher in /kl/ and /pl/ than in /l/ in intervocalic position, due to its role in achieving the velar closure target of /k/ or the labial closure target of /p/. Consequently, when the tongue-tip gesture for /l/ is initiated, the tongue tip is in a different state – i.e. closer to the palate – in the /kl/ and /pl/ environments than in the /l/ environment. This results in a decrease in the amount of time it takes for the tongue tip to reach its target for /l/, as illustrated in Fig. 5b. This coarticulatory effect, which we call BIOMECHANICAL SHORTENING, leads to a non-symmetrical pattern of target achievement relative to the vowel, since the duration of the gestural activation interval for /l/ is modified.

2.3 The relation between empirical measurements and model predictions

The gestural timing patterns which are most directly predicted by the Articulatory Phonology coupled oscillators model are almost always

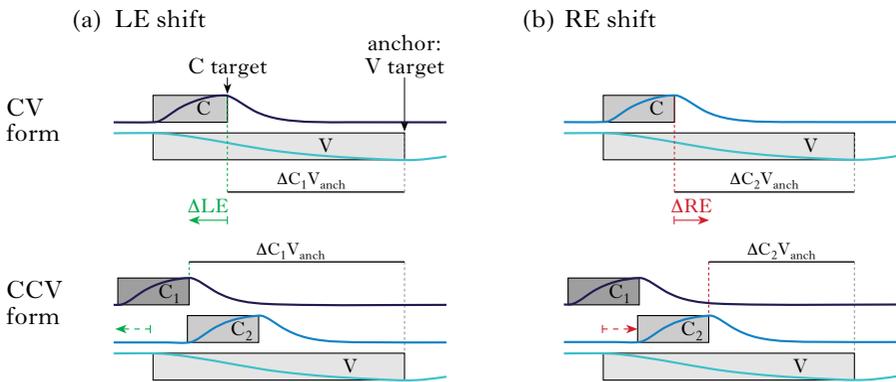


Figure 6

Estimation of C-centre effect for complex onset coordination: (a) left-edge shift (ΔLE) = ($\Delta C_1 V$ in the CV form – $\Delta C_1 V$ in the CCV form); (b) right-edge shift (ΔRE) = ($\Delta C_2 V$ in the CV form – $\Delta C_2 V$ in the CCV form).

measured indirectly. This holds in our approach as well. To understand the indirect character of this measurement, it is important to clarify several points, which are discussed in relation to the schematic representations of empirical data in Fig. 6. The figure shows consonantal and vocalic gestural activation intervals for a CV form and a CCV form, along with hypothetical movement trajectories generated by the gestures.

First, although the coupled oscillators model generates a pattern of INITIATION of gestural activation, approaches to empirical measurement of this pattern are derived from the timing of gestural target achievement. The reason for this is that gestural target achievements for consonantal constrictions are relatively easy to locate in articulatory data; in contrast, attempts to measure the timing of gestural initiation are frequently confounded by interactions with preceding articulatory postures and by effects of coarticulation. Second, the interval which is used to assess timing patterns is the duration between consonantal target achievement and some later event, such as the achievement of a vocalic target. These intervals are labelled $\Delta C_1 V_{\text{anch}}$ and $\Delta C_2 V_{\text{anch}}$ in Fig. 6. Third, the C-centre effect is always calculated by a comparison, in particular, a comparison of $\Delta C_1 V_{\text{anch}}$ and $\Delta C_2 V_{\text{anch}}$ in singleton environments (i.e. /l/, /k/ or /p/) to $\Delta C_1 V_{\text{anch}}$ and $\Delta C_2 V_{\text{anch}}$ in cluster environments (i.e. /kl/ or /pl/). Specifically, the comparison derives ‘shift measures’. The left-edge shift, ΔLE , is defined as $\Delta C_1 V_{\text{anch}}$ in the singleton context minus $\Delta C_1 V_{\text{anch}}$ in the cluster context. Similarly, the right-edge shift, ΔRE , is defined as $\Delta C_2 V_{\text{anch}}$ in the singleton context minus $\Delta C_2 V_{\text{anch}}$ in the cluster context. These between-environment differences in $\Delta C_1 V_{\text{anch}}$ and $\Delta C_2 V_{\text{anch}}$ are labelled with solid arrows in Fig. 6.

Finally, there are two important assumptions that underlie all approaches to evaluating the coupled oscillators model. First, it is assumed that in a singleton environment, a consonantal constriction gesture is initiated at approximately the same time as the vocalic constriction gesture. This is predicted by the coupled oscillators model, and there is a substantial body of literature supporting it (Browman & Goldstein 2000, Nam *et al.* 2009, Marin & Pouplier 2010, Shaw *et al.* 2011, Hermes *et al.* 2013, Gafos *et al.* 2014). Second, it is assumed that the duration from the initiation of gestural activation to achievement of target does not differ between singleton and cluster contexts. This assumption of constant onset-to-target duration is necessary to test predictions about the timing of gestural initiation – which is what the coupled oscillators model generates – using observations of the timing of target achievement. Changes in the timing of initiation are represented by the broken arrows in Fig. 6.

The empirical data examined in the following sections thus consist of pairs of values, left-edge shift (ΔLE) and right-edge shift (ΔRE). These were calculated using the method described above on a by-speaker by-condition by-target basis from the averages of CV_{anch} intervals in CV and CCV environments.

3 Case studies on variability in syllable coordination

The phonological aspects of syllable coordination in Articulatory Phonology are the underlying syllable parse (simplex or complex) and the coupling topologies and coupling forces. In two case studies, we investigate the C-centre effect as a phonetic reflex of the underlying coupling structures. We will show that the C-centre pattern is not always congruent with the underlying coupling structure, and how this problem can be solved in dynamical systems theory. We test whether the coupled oscillators model can handle a high amount of variability when the linking model is sufficiently elaborated. The first dataset is concerned with ageing data (younger *vs.* older speakers; §3.1) and the second one with pathological data (Essential Tremor patients treated with deep brain stimulation *vs.* healthy control speakers; §3.2). In both studies, we investigate syllable-coordination patterns for two complex onsets in German, /pl/ and /kl/, on the basis of the C-centre paradigm. We compare different models by testing the complex onset parse with balanced and imbalanced coupling strength, as well as with and without biomechanical correction. In addition, we run a model that changes the coupling structure itself, i.e. switching from simplex to complex onset parse, in order to account for data variability in the different datasets. The models are described as in (4).

(4) a. *Heterogeneous unconstrained model*

We allow for the phonological syllable parse (simplex or complex onset) to change for each data point.

b. *Complex balanced model*

We assume a complex onset parse for all datasets with fixed parameters for coupling strength (this is how the coupled oscillators model is often implemented).

c. *Complex imbalanced model*

We assume the complex onset parse for all datasets, but allow the coupling-strength parameter to vary, in order to generate asymmetrical shift patterns (i.e. our first model extension).

d. *Complex imbalanced model with biomechanical correction*

We adjust the biomechanical correction parameter to reflect segmentally conditioned variation in the data (i.e. our second model extension).

3.1 Case study 1: ageing

Ageing entails several physiological changes which can lead to deficits in movement and posture, involving not only limbs and torso, but also the organs used in speech. Studies on non-speech motor control reveal that movements are slowed down in older populations (Cooke *et al.* 1989, Seidler *et al.* 2002). The slowing-down process involves changes in the timing of movement patterns. Furthermore, changes in the timing of the movement components have been reported which involve an asymmetry between the acceleration and deceleration phases, revealing longer deceleration phases (Cooke *et al.* 1989, Ketcham & Stelmach 2004).

There is also evidence that age affects the precision of speech motor control. In a study on German, Hermes *et al.* (2018) found effects in speech similar to those reported for general motor control. Using 3D electromagnetic articulography, they tracked the movements of the lips and the tongue during the production of consonants and vowels in natural sentences, and found a slowing-down of articulatory movements that was accompanied by a change in the intragesural timing patterns of the primary constrictors during consonant and vowel production, revealing an asymmetry between acceleration and deceleration phases in the way that the deceleration phases were prolonged.

Based upon the finding reported in Hermes *et al.* (2018), we assume that ageing also affects the timing *between* gestures, i.e. it leaves a signature in the outcome of syllable-internal coordination patterns. We therefore investigate variability of the complex onsets /pl/ and /kl/, comparing older and younger German speakers.

3.1.1 *Method.* The dataset on ageing is based on recordings from Hermes *et al.* (2018). It consists of five older speakers, aged 70–80, and five younger speakers, aged 20–30. The articulatory recordings were

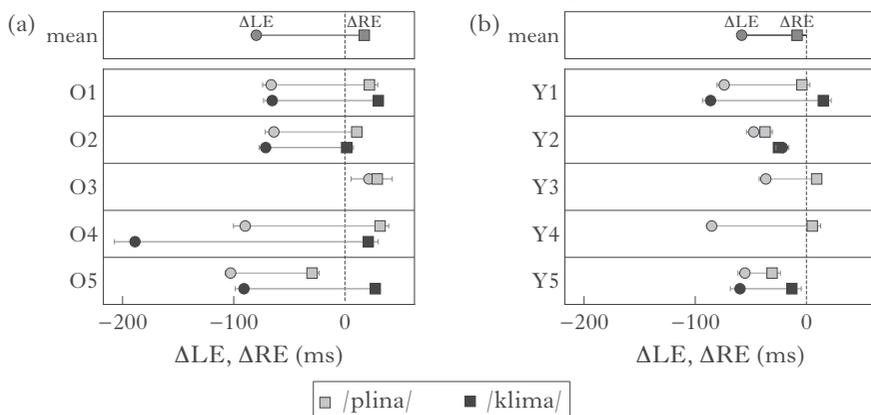


Figure 7

Empirical Δ LE and Δ RE for (a) older and (b) younger speakers in the ageing dataset. The vertical dashed lines mark the point in time where the respective shifts for C_2 (Δ RE) and C_1 (Δ LE) amount to 0 ms (no shift). Positive values indicate a rightward shift towards the V in complex onset patterns (Δ RE; squares) and negative values indicate a shift away from the V (Δ LE; circles).

carried out with a 3D Carstens Electromagnetic Articulograph (AG501) at the IfL Phonetics department in Cologne. To track the movements of the articulators, sensors were placed on the upper and lower lips, and the tongue tip, blade and dorsum. The kinematic data were recorded at 1200 Hz, downsampled to 250 Hz and smoothed with a three-step floating mean. The speech material contained disyllabic target words bearing the nuclear accent in a carrier sentence: e.g. *Er hat wieder /klima/ gesagt* ‘He said ‘climate’ again’. Every C-centre measure needs a triple of target words containing the structure C_1V , C_2V , C_1C_2V , as in (5).

- | | | | | | | |
|------------|--------------|---------|---------------|--------------|---------|-----------|
| (5) C_1V | <i>Pina</i> | /pina/ | (girl’s name) | <i>Kina</i> | /kina/ | (nonce) |
| C_2V | <i>Lima</i> | /lima/ | (place name) | <i>Lima</i> | /lima/ | |
| C_1C_2V | <i>Plina</i> | /plina/ | (nonce) | <i>Klima</i> | /klima/ | ‘climate’ |

The articulatory data were annotated using the EMU Speech Database System (Cassidy & Harrington 2001). Landmarks in the articulatory domain for consonantal and vocalic gestures were identified in the vertical plane. The onsets and targets (local minima and maxima) of the respective gestures were labelled using zero-velocity crossing in the velocity curve. The C-centre measures, left-edge shift and right-edge shift were computed as described in §2.3.

3.1.2 *Results. Empirical data:* LE and RE shifts were calculated for older and younger speakers (O1–5 and Y1–5 respectively) for the two cluster types in Fig. 7, /pl/ (grey) and /kl/ (black). Positive RE shift values indicate

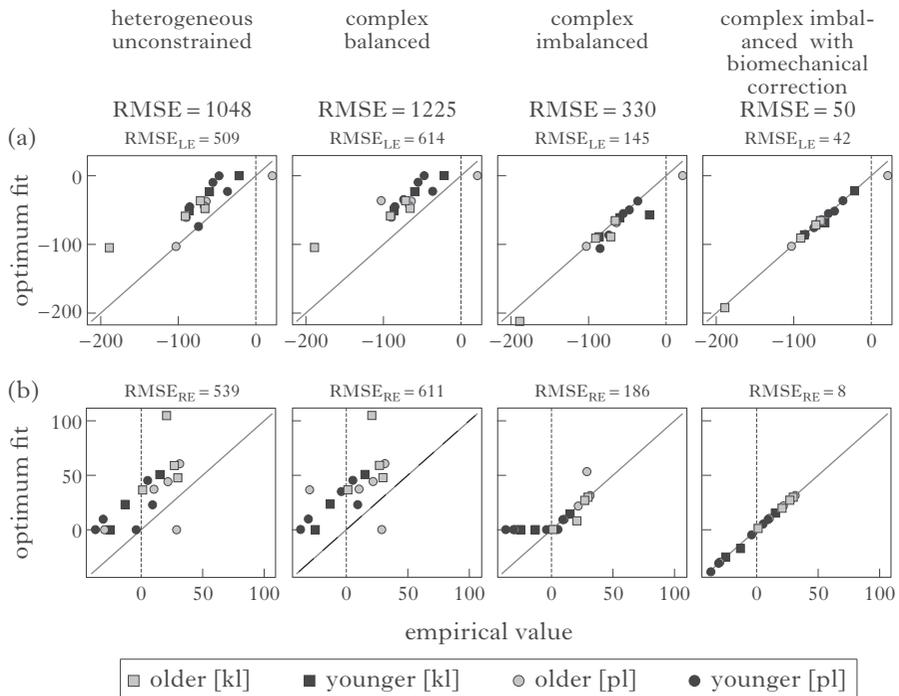


Figure 8

Assessment of model fits for ageing dataset: (a) ΔLE ; (b) ΔRE . The x-axis shows the empirical value of the temporal shift of the consonantal gesture(s); the y-axis shows the optimal model-generated values. The RMSE of the correlation between empirical data and model fits are shown in each panel.

a shift towards the V in cluster environment, revealing a higher degree of overlap compared to simplex onsets. Negative values of LE shift indicate a shift away from the V, corresponding to a lower degree of overlap between C and V. The vertical dashed line is the vowel onset. The prototypical C-centre pattern predicts a symmetrical shift for C_1 (circles) and C_2 (squares), where the midpoints of the horizontal bars are centred on the vertical dashed line – like a seesaw where the midpoint of the board is located at a pivot point (here 0 ms). However, this symmetry is clearly not observed in our data.

In the younger speaker group in (b), the C_1 shift amounts on average to -59 ms for /k/ and -60 ms for /p/. But there is no corresponding pattern for the C_2 shift. Indeed, the C_2 shifts are rather small, and for some of the speakers they go in the wrong direction (-5 ms for /k/ and -11 ms for /p/ across all younger speakers). This type of surface pattern reveals no evidence for a C-centre organisation from a conventional perspective, since the overlap between the prevocalic C and the following V does not increase when a consonant is added to the syllable. A different picture arises when

we look at the shift patterns for the older speakers. In this group, the C_1 shift is on average 104 ms for /kl/ and 60 ms for /pl/. There is also a C_2 shift for the prevocalic C towards the following V (22 ms in /kl/ and 12 ms in /pl/ across all older speakers), but the C_2 shift is smaller than the C_1 shift.

Modelling data: We now discuss the data generated in the computational model for the different model implementations described in §2. We evaluate how well the models can generate the shifts of C_1 (LE fit) and C_2 (RE fit) observed in the empirical data. Table II shows the root mean square error (RMSE) of LE fit and RE fit, and the total RMSE for all models. RMSE is used to quantify the difference between the optimal values generated by the computational model and the observed values in empirical data. The lower the RMSE values, the better the fit. Figure 8 shows the relations between empirical values and model predictions for a subset of the models tested. The models can be usefully compared in terms of the RMSE of the fit between empirical data (x -axis) and the optimal model-generated values (y -axis), as well as the total RMSE (see Table II). The model fits for /kl/ and /pl/ for younger and older speakers are shown in each panel by the black and grey circles and squares.

			RMSE		
	topology	coupling	LE	RE	total
without bio-mechanical correction	complex	imbalanced	145	186	331
	complex	balanced	614	611	1225
	heterogeneous unconstrained		509	539	1048
	heterogeneous constrained		764	483	1246
	simplex		1226	340	1566
with bio-mechanical correction	complex	imbalanced	42	8	50
	complex	balanced	297	304	601
	heterogeneous unconstrained		207	223	430
	heterogeneous constrained		490	217	707
	simplex		960	205	1165

Table II

Model performance for the ageing dataset. The lower the values, the better the fit.

As expected, the best-performing model is the complex imbalanced model with biomechanical correction, which fits the data nearly perfectly. In all cases, the models with biomechanical correction outperform their counterparts without biomechanical correction. This is due to the fact that these models have an additional parameter, which, as argued above, is consistent with known biomechanical interactions. Furthermore, within any given set of models which do or do not have biomechanical

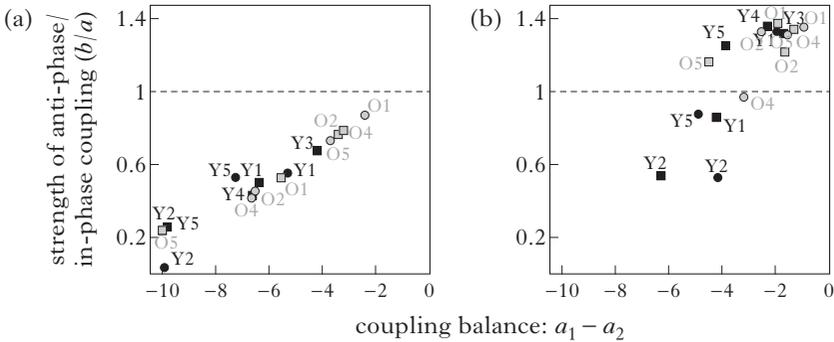


Figure 9

Optimised coupling balance ($a_1 - a_2$) and the strength of anti-phase coupling relative to in-phase coupling (b/a) for the extended models: (a) imbalanced coupling; (b) imbalanced coupling with biomechanical correction. The x -axis shows the coupling balance ($a_1 - a_2$); a more negative number indicates a greater degree of imbalance, such that C_2 is more strongly coupled to V than C_1 . The y -axis shows the strength of anti-phase coupling relative to in-phase coupling (b/a); a value of 1 corresponds to equally strong in-phase and anti-phase coupling. (Note that O3 is excluded, because of the poor-quality fit.)

correction, we observe that the complex imbalanced model outperforms all other models. The heterogeneous unconstrained model outperforms the complex balanced and simplex balanced models (see Table II), which is expected, because it selects whichever of these best fits the empirical data. Notably, the heterogeneous constrained model (we assume a simplex onset parse for the young speakers and a complex onset parse for the older speakers) performs worse than the traditional complex balanced model for all speakers, but better than the simple model, again for all speakers.

Figure 9 shows optimised parameters for the two extended models: (a) imbalanced coupling and (b) imbalanced coupling with biomechanical correction. The y -axis shows the strength of anti-phase coupling relative to in-phase coupling, and the x -axis the coupling balance, i.e. the strength of C_1 - V coupling minus the strength of C_2 - V coupling. In both cases, we find a more balanced coupling strength for older speakers than for younger speakers. In a balanced coupling structure, the value for the in-phase coupling modes ($a_1 - a_2$) should amount to zero, while more negative numbers indicate a greater imbalance, in that C_2 is more strongly coupled to V than C_1 is to V . For the model parameter which corresponds to the relative strength of anti-phase and in-phase coupling (b/a), a value of 1 corresponds to equally strong in-phase and anti-phase coupling (a is the average of a_1 and a_2).

The figure shows that older speakers tend to have a more balanced coupling than younger speakers (Fig. 9a), and when we add a biomechanical correction (Fig. 9b) the older speakers reveal a stronger anti-phase coupling relative to in-phase coupling than the younger speakers.

Importantly, in the above analysis, both speaker groups use the same coupling structure (i.e. the same phonological syllable parse), but the coupling strengths differ, leading to more symmetrical shifts for the older than for the younger speakers. Our modelling thus shows that the asymmetrical pattern in younger speakers need not be interpreted as evidence for a structural change in phonological syllable parse.

3.2 Case study 2: pathological speech

Essential Tremor is one of the most common movement disorders (Deuschl & Elble 2009), and is characterised by an action tremor affecting limbs or other body parts. A very successful treatment has been developed for patients who are medication-resistant. This treatment is DEEP BRAIN STIMULATION (DBS), where a medical device sends electrical impulses through implanted electrodes to specific parts of the brain, in order to suppress the tremor. The target region for the electrode implantation is the thalamus, more specifically the *nucleus ventralis intermedius*; Flora *et al.* 2010). Although there is suppression of the tremor, patients report that stimulation has a deleterious effect on their speech (e.g. slurred speech sounds, less flexible tongue movements, shortness of breath).

In fast-syllable repetition tasks, Mücke *et al.* (2014) and Mücke *et al.* (2018) show that Essential Tremor patients treated with DBS have coordination problems of vocal tract movements in terms of imprecise consonant articulation (spirantisation of stop consonants) and slowness. Speech deteriorates under stimulation, but it was not clear from the neuro-anatomical data whether this was due to an affection of the upper motor fibres of the internal capsule caused by the current spread of the activated electrode or to the aggravation of pre-existing cerebellar deficits, or to a combination of the two (Hermes *et al.* 2019).

3.2.1 Method. The dataset on DBS is based on recordings from Mücke *et al.* (2018) and Hermes *et al.* (2019) from nine Essential Tremor patients aged between 31 and 73 and nine age- and gender-matched control speakers. All Essential Tremor patients had had surgery (the DBS implantation) at least four months prior to their participation in the study. The surgery helped to suppress the tremor for all patients, but as an inadvertent side-effect the speech worsened, especially when the stimulation was activated (Mücke *et al.* 2018). The patients were recorded with activated (DBS-ON) and inactivated (DBS-OFF) stimulation within one recording session. For both measurements, the sensors remained at the articulators for both measurements (DBS-ON and DBS-OFF). The order of the stimulation (DBS-ON and DBS-OFF) was randomised, and before each testing the stimulation settings were maintained for at least 20 minutes. All procedures with respect to recordings, data processing, annotations and measures were comparable with those adopted for the ageing dataset (see §3.1.1), and the speech material for the /kl/ and /pl/ clusters corresponded to the speech material in §3.1.1 for the ageing dataset.

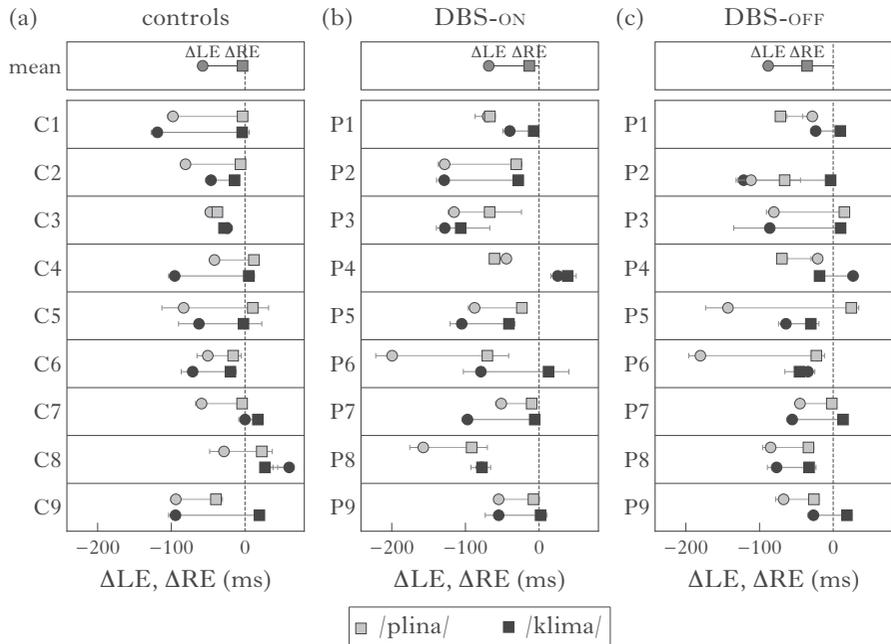


Figure 10

Empirical ΔLE and ΔRE for the DBS dataset. The vertical dashed lines mark the point in time where the respective shifts for C_2 (ΔRE) and C_1 (ΔLE) amount to 0 ms (no shift). Positive values indicate a rightward shift towards the V in complex onset patterns (ΔRE ; squares) and negative values indicate a shift away from the V (ΔLE ; circles).

3.2.2 Results. Empirical data: We compared syllable-coordination patterns in Essential Tremor patients with activated (DBS-ON) and inactivated stimulation (DBS-OFF) with age- and gender-matched healthy controls. We computed LE and RE shifts in the same way as for the ageing dataset. Figure 10 shows the shifts for /p/ and /k/ for (a) the controls, (b) patients in the DBS-ON group and (c) patients in the DBS-OFF group. Positive values of ΔRE indicate a delay of C_2 onset relative to vowel onset, and negative values of ΔLE indicate an earlier C_1 onset relative to vowel onset.

In the controls, there is effectively no shift for C_2 (average C_2 shifts in controls: 1 ms for /k/ and -6 ms for /p/), while the C_1 shifts amount to -50 ms for /k/ and -64 ms for /p/. This pattern is not in line with the standard coupled oscillators model, which predicts that the onset of C_2 will be delayed relative to the vowel onset. In the patient groups, the patterns deviate even further from the prediction. With stimulation inactivated (DBS-OFF), there is a shift of C_2 , but it goes in the wrong direction: C_2 shifts *away* from the following V, instead of towards it.

The overlap between C₂ and the following V decreases when a C is added to the beginning of the word (C₂ shift in DBS-OFF: -9 ms for /kl/ and -28 ms for /pl/; C₁ shift in DBS-OFF: -52 ms for /kl/ and -85 ms for /pl/). The pattern worsens under activated stimulation. In the DBS-ON condition, the shift for C₂ away from the following V *increases*, revealing a strong shift of C₂ in the wrong direction (C₂ shift in DBS-ON: -24 ms for /kl/ and -48 ms for /pl/; C₁ shift in DBS-ON: -76 ms for /kl/ and -101 ms for /pl/).

Modelling data: Analogously to the ageing data, we consider how well variations of computational model fit the data. Results are shown in Table III in terms of RMSE values. As for the ageing data, the model optimisations show that the complex imbalanced model with biomechanical correction fits the data for the pathological speech and the controls best. The heterogeneous unconstrained model (adjusting complex/simplex onset parse with respect to each data point) provides a better fit than either the traditional simple or the complex balanced model; this again is expected, because it selects the better-fitting model for each data point. Interestingly, the simple model provides a better fit than either the heterogeneous constrained or complex model for the DBS dataset. This might lead to the interpretation that DBS patients employed a simple coordination pattern. However, because the complex imbalanced model provides a better fit than either of the simplex or heterogeneous unconstrained models, we conclude that DBS patients do employ a complex parse.

			RMSE		
	topology	coupling	LE	RE	total
without bio-mechanical correction	complex	imbalanced	680	1390	2070
	complex	balanced	2595	2680	5275
	heterogeneous unconstrained		879	1869	2748
	heterogeneous constrained		1745	1898	3643
	simplex		1839	1551	3390
with bio-mechanical correction	complex	imbalanced	530	390	919
	complex	balanced	1696	1664	3360
	heterogeneous unconstrained		411	677	1088
	heterogeneous constrained		1413	844	2257
	simplex		1542	627	2169

Table III

Model performance for the DBS dataset. The lower the values, the better the fit.

Figure 11 shows the relations between empirical values and model predictions for a selection of models: if we assume that the patients have

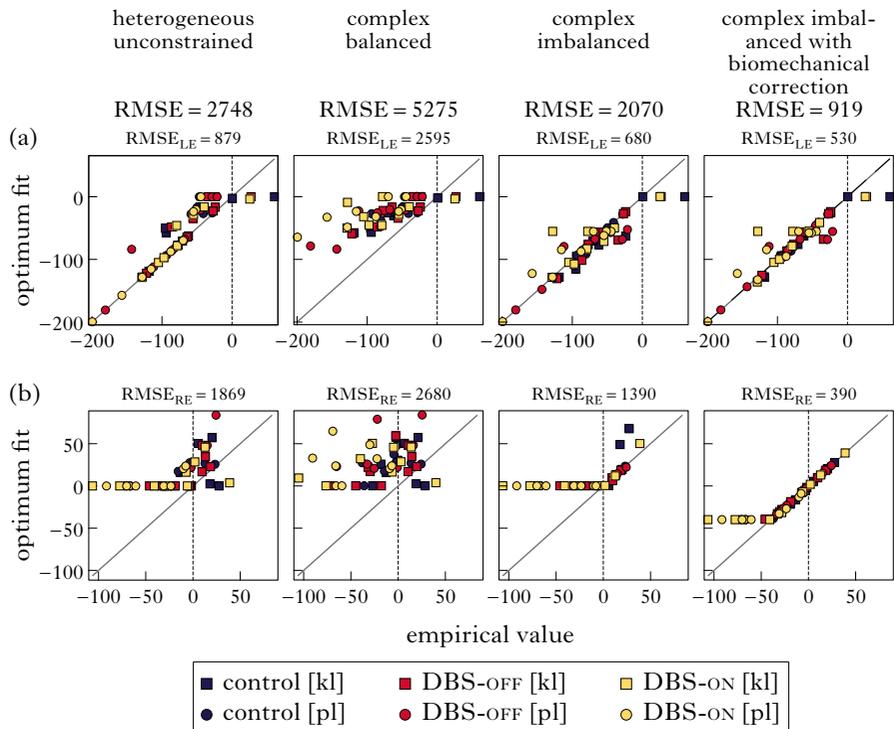


Figure 11

Assessment of model fits for the DBS dataset: (a) ΔLE ; (b) ΔRE . The x -axis shows the empirical value of the temporal shift of the consonantal gesture(s); the y -axis shows the optimal model-generated values. The RMSE of the correlation between empirical data and model fits are shown in each panel.

simplex onsets and controls complex onsets (i.e. the heterogeneous unconstrained model), the fit is even worse than if we assume that all have a simplex onset parse. As expected, the complex imbalanced model outperforms all other models. Given phonological evidence that German has complex onsets in these environments, the simplest account is one which is consistent with this evidence. This supports the interpretation that DBS patients and controls employed a complex syllable parse, with parametric variation in coupling strength.

Regarding the parametric variation that is observed in model fits, our results show that DBS patients exhibited significantly less balanced coupling ($a_1 - a_2$) than controls (control: mean = -2.1 , SD = 4.4 ; Essential Tremor patients: mean = -5.1 , SD = 4.0 , $p = 0.015$, $t(52) = 2.52$). Furthermore, DBS patients exhibit a weaker strength of anti-phase coupling relative to in-phase coupling (b/a) in the OFF condition than in the ON condition (DBS-OFF: mean = 1.0 , SD = 0.6 ; DBS-ON: mean = 1.3 , SD = 0.3 , $p = 0.027$, $t(17) = -2.41$). These two mechanisms can be interpreted as a massive

weakening of the anti-phase mode for DBS patients (cf. Hermes *et al.* 2019), leading to a pattern on the surface that could be misinterpreted as a simplex onset organisation (due to the missing rightward shift).

4 Discussion

4.1 Variation in the ageing dataset

The comparison of articulatory timing patterns in German has revealed differences between younger and older speakers. While the older speaker group exhibited the predicted rightwards shift for the prevocalic C in complex onsets, the younger speakers did not. For both groups, the shift patterns for C₁ and C₂ were asymmetrical in stop–lateral sequences.

The empirical data for the younger speaker group might be taken as evidence against a complex syllable parse. The younger speakers showed no rightward shift at all; this is also reported by Brunner *et al.* (2014) for /pl/ in German. In complex onsets, the prevocalic C is expected to shift towards the following V to make room for the added C in C₁C₂V sequences. Only in simplex onsets should no rightmost shift of the prevocalic C occur, reflecting the fact that the added C is not part of the syllable, and therefore does not interfere with the syllable-internal organisation. Furthermore, the pattern for the older speakers can, at least to some extent, also be interpreted as a problem for the prediction of complex onset parses, since the shift of C₂ was rather small, and much smaller than the LE shift. With a standard implementation of the coupled oscillators model in which coupling is balanced, a symmetrical shift pattern between C₁ and C₂ is expected (Browman & Goldstein 2000, Nam *et al.* 2009, Gafos *et al.* 2010, Marin & Pouplier 2010, Shaw *et al.* 2011, Hermes *et al.* 2013, Shaw & Gafos 2015, Hermes *et al.* 2017).

However, our modelling of the empirical data shows that a complex onset parse need not be rejected if the linking model is elaborated to allow for gradient variation in coupling strength. Indeed, the best model fit for both younger and older speakers was observed when coupling strength was allowed to vary gradually (the complex imbalanced model). Adding the biomechanical correction parameter further enhanced the performance of the model.

Impressionistically, it has been observed that older speakers tend to hyperarticulate. This is consistent with our finding that older speakers have a higher strength of anti-phase coupling relative to in-phase coupling than younger speakers, as well as more balanced coupling. We assume that a greater degree of balance in coupling strength incurs a higher cost for the physical control system in producing more prototypical, canonical forms for complex onsets. In contrast, younger speakers showed tendencies to hypoarticulate (low-cost behaviour; imbalanced couplings leading to asymmetrical shifts). This interpretation is further supported by the observation that older speakers had lower velocities for C₂ (/l/) in C₁C₂V clusters than younger speakers.

4.2 Variation in the pathological dataset

Our empirical data revealed systematic differences in the syllable-coordination patterns between Essential Tremor patients and control speakers, as well as between the Essential Tremor patients with activated and inactivated stimulation. The control speakers showed no shift for C_2 towards the following V, even though this shift is predicted for the complex onset parse. This is incongruent with the predictions of the standard coupled oscillators model, but, as with the ageing data, we have shown that we can account for this variability in a complex coupling structure by allowing for an imbalanced parameterisation of coupling strength.

However, the LE and RE patterns for the patients were different from the controls, which is potentially challenging for our assumptions. The patients showed a shift of C_2 , but the shift was in the wrong direction: earlier, rather than later. The phonological syllable parse predicts an increase in overlap between C_2 and the following V, but we found a decrease in overlap in the patients' data. This atypical pattern was magnified under activated stimulation: in the DBS-ON condition the shift of C_2 away from the V considerably increased.

From the empirical data, we can conclude that Essential Tremor patients treated with DBS show a deterioration in syllable production in the condition with inactivated stimulations. The syllable patterns show inefficient timing, since the prevocalic C shifts away from the following V when a consonant is added to the syllable. The timing problem in the DBS-OFF condition is likely due to pre-existing cerebellar deficits of the disease, which can be interpreted as signs of dysarthria (Kronenbueger *et al.* 2009). The problems worsened under stimulation, inducing stronger shifts of C_2 in the wrong direction. It is likely that the spread of current from the electrodes implanted in the specific target area of the brain induced this type of deviant timing pattern (Mücke *et al.* 2018, Hermes *et al.* 2019).

5 Conclusion

We have shown that stereotyped interpretations of phonetic measures can lead to the types of error in phonological interpretation in (6).

(6) a. *Revision of the theory*

The theory is adapted in an ad hoc manner to generate surface variation. This often results in a proliferation of categories or an incorrect association of surface patterns with categories.

b. *Rejection of the theory*

The underlying theory has to be rejected, since its predictions are violated by the phonetic output.

In the context of the C-centre phenomenon, the error in (6a) would involve associating simplex and complex onset parses with different speaker groups, or with different speakers or targets in different conditions. Specifically, we could infer that older speakers use complex onsets, while younger speakers use simplex onsets. For pathological speech, we might conclude that Essential Tremor patients change from a complex onset parse to a simplex parse. However, we argue that such interpretations are misguided, and result from an overly simplistic model of the link between phonological knowledge (a coupling topology and the respective coupling strengths) and articulatory timing in the phonetic output. We showed that, with a more sophisticated model, there is no need to reject the underlying phonological theory that word-initial clusters in German have a complex syllable topology. We propose that the same conclusions and analytical considerations are likely to be applicable in other languages as well.

In the case of (6b), the observed variability in gestural timing patterns could be interpreted as evidence against underlying phonological theory. The observed absence of consistent symmetric LE and RE shifts in the data might argue against the coupled oscillators model, and perhaps one could speculate that C-centre measures merely reflect segmental variation in overlap patterns. Again, we feel this would be an incorrect conclusion, because it relies on an overly simplistic model.

Ultimately, we advocate more careful approaches to assessing phonological theories with empirical phonetic data. Such approaches always depend on a linking model, which may be more or less explicit, and more or less sophisticated. In the case of the Articulatory Phonology coupled oscillators model, the problem is not explicitness, but rather that artificial constraints are imposed on the model. Relaxing these constraints increases the power of the model, and clarifies interpretation of empirical data. Part of the problem may arise from a tendency to routinise empirical testing procedures – i.e. the stereotyped use of phonetic tools – with uncritical reliance on assumptions which may not be justified.

In sum, we emphasise that testing theoretical predictions in the domain of speech is almost never a black-and-white question of confirming or disconfirming a hypothesis. It is rarely the case that empirical observations straightforwardly determine the fate of our theoretical models. Instead, there are often multiple layers of interacting mechanisms which complicate the relations between theory and observation. The phonetic surface representation can be strongly masked by various factors in the multidimensional speech system, and we want to draw attention to the fact that incongruencies between phonetic analysis and phonological theory can help us to understand the underlying structural components interacting on different layers. We must always consider the possibility that a hypothesis is correct, but nonetheless fails to match the world, for reasons that remain to be established. A shift in perspective can explain surface differences, without necessitating rejection or substantial revision of the model.

REFERENCES

- Abercrombie, David (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Anderson, Stephen R. (1981). Why phonology isn't 'natural'. *LI* 12. 493–539.
- Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *JPh* 26. 3–25.
- Arvaniti, Amalia (2009). Rhythm, timing and the timing of rhythm. *Phonetica* 66. 46–63.
- Arvaniti, Amalia & Tara Rodriguez (2013). The role of rhythm class, speaking rate, and *F0* in language discrimination. *Laboratory Phonology* 4. 7–38.
- Bolozky, Shmuel (1997). Israeli Hebrew phonology. In A. S. Kaye (ed.) *Phonologies of Asia and Africa (including the Caucasus)*. Vol. 1. Winona Lake, Ind.: Eisenbrauns. 287–311.
- Blumstein, Sheila E. (1991). The relation between phonetics and phonology. *Phonetica* 48. 108–119.
- Browman, Catherine P. & Louis Goldstein (1986). Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.
- Browman, Catherine P. & Louis Goldstein (1988). Some notes on syllable structure in articulatory phonology. *Phonetica* 45. 140–155.
- Browman, Catherine P. & Louis Goldstein (1992). Articulatory phonology: an overview. *Phonetica* 49. 155–180.
- Browman, Catherine P. & Louis Goldstein (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5. 25–34.
- Brunner, Jana, Christian Geng, Stavroula Sotiropoulou & Adamantios Gafos (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology* 5. 403–454.
- Cassidy, Steve & Jonathan Harrington (2001). Multi-level annotation in the Emu speech database management system. *Speech Communication* 33. 61–77.
- Chang, Woohyeok (2012). On the relation between phonetics and phonology. *Linguistic Research* 29. 127–156.
- Chitoran, Ioana & Abigail C. Cohn (2009). Complexity in phonetics and phonology, gradience, categoriality, and naturalness. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.) *Approaches to phonological complexity*. Berlin & New York: Mouton de Gruyter. 21–46.
- Cho, Taehong (2006). Manifestation of prosodic structure in articulatory variation: evidence from lip kinematics in English. In Louis Goldstein, Douglas Whalen & Catherine T. Best (eds.) *Papers in laboratory phonology 8*. Berlin & New York: Mouton de Gruyter. 519–548.
- Cooke, J. D., S. H. Brown & D. A. Cunningham (1989). Kinematics of arm movements in elderly humans. *Neurobiology of Aging* 10. 159–165.
- Deuschl, Günther & Rodger Elble (2009). Essential tremor – neurodegenerative or nondegenerative disease towards a working definition of ET. *Movement Disorders* 24. 2033–2041.
- D'Imperio, Mariapaola, Caterina Petrone & Noël Nguyen (2007). Effects of tonal alignment on lexical identification. In Carlos Gussenhoven & Tomas Riad (eds.) *Tones and tunes*. Vol. 2: *Experimental studies in word and sentence prosody*. Berlin & New York: Mouton de Gruyter. 79–106.
- Flora, Eliana Della, Caryn L. Perera, Alun L. Cameron & Guy J. Maddern (2010). Deep brain stimulation for essential tremor: a systematic review. *Movement Disorders* 25. 1550–1559.

- Fowler, C. A., P. Rubin, R. E. Remez & M. T. Turvey (1980). Implications for speech production of a general theory of action. In B. Butterworth (ed.) *Language production*. Vol. 1: *Speech and talk*. London: Academic Press. 373–420.
- Gafos, Adamantios I. & Stefan Benus (2006). Dynamics of phonological cognition. *Cognitive Science* **30**. 905–943.
- Gafos, Adamantios I., Philip Hoole, Kevin Roon & Chakir Zeroual (2010). Variation in overlap and phonological grammar in Moroccan Arabic clusters. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.) *Laboratory phonology 10*. Berlin & New York: De Gruyter Mouton. 657–698.
- Gafos, Adamantios I., Simon Charlow, Jason A. Shaw & Philip Hoole (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *JPh* **44**. 152–166.
- Gafos, Adamantios I., Jens Roeser, Stavroula Sotiropoulou, Philip Hoole & Chakir Zeroual (2020). Structure in mind, structure in vocal tract. *NLLT* **38**. 43–75.
- Goldstein, Louis, Dani Byrd & Elliot Saltzman (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.) *Action to language via the mirror neuron system*. Cambridge: Cambridge University Press. 215–249.
- Goldstein, Louis, Hosung Nam, Elliot Saltzman & Ioana Chitoran (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki & J. Shen (eds.) *Frontiers in phonetics and speech science: Festschrift for Wu Zongji*. Beijing: Commercial Press. 239–249.
- Hawkins, Sarah (1992). An introduction to task dynamics. In Gerard J. Docherty & D. Robert Ladd (eds.) *Papers in laboratory phonology II: gesture, segment, prosody*. Cambridge: Cambridge University Press. 9–25.
- Hermes, Anne, Doris Mücke & Martine Grice (2013). Gestural coordination of Italian word-initial clusters: the case of 'impure s'. *Phonology* **30**. 1–25.
- Hermes, Anne, Doris Mücke & Bastian Auris (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *JPh* **64**. 127–144.
- Hermes, Anne, Jane Mertens & Doris Mücke (2018). Age-related effects on sensorimotor control of speech production. In *Proceedings of Interspeech 2018*. 1526–1530. Available (January 2020) at https://www.isca-speech.org/archive/Interspeech_2018/pdfs/1233.pdf.
- Hermes, Anne, Doris Mücke, Tabea Thies & Michael T. Barbe (2019). Coordination patterns in Essential Tremor patients with Deep Brain Stimulation: syllables with low and high complexity. *Laboratory Phonology* **10**(1):6. <http://doi.org/10.5334/labphon.141>.
- Honorof, Douglas N. & Catherine P. Browman (1995). The center or edge: how are consonant clusters organized with respect to the vowel? In Kjell Elenius & Peter Branderud (eds.) *Proceedings of the 13th International Congress of the Phonetic Sciences*. Vol. 4. Stockholm: KTH & Stockholm University. 552–555.
- Hyman, Larry M. (1975). *Phonology: theory and analysis*. New York: Holt, Rinehart & Winston.
- Keating, Patricia A. (1988). The phonology–phonetics interface. In Frederick J. Newmeyer (ed.) *Linguistics: the Cambridge survey*. Vol. 1: *Grammatical theory*. Cambridge: Cambridge University Press. 281–302.
- Kelso, J. A. Scott (1995). *Dynamic patterns: the self-organization of brain and behavior*. Cambridge, Mass.: MIT Press.
- Ketcham, Caroline J. & George E. Stelmach (2004). Movement control in the older adult. In Richard W. Pew & Susan B. Van Hemel (eds.) *Technology for adaptive aging*. Washington, D.C.: National Academies Press. 64–92.
- Krivokapić, Jelena (2013). Rhythm and convergence between speakers of American and Indian English. *Laboratory Phonology* **4**. 39–65.

- Kronenbuerger, Martin, Jürgen Konczak, Wolfram Ziegler, Paul Buderath, Benedikt Frank, Volker A. Coenen, Karl Kiening, Peter Reinacher, Johannes Noth & Dagmar Timmann (2009). Balance and motor speech impairment in essential tremor. *Cerebellum* **8**. 389–398.
- Kühnert, Barbara, Philip Hoole & Christine Mooshammer (2006). Gestural overlap and C-center in selected French consonant clusters. *Proceedings of the 7th International Seminar on Speech Production, Ubatuba, Brazil*. 327–334.
- Ladd, D. Robert (2006). Segmental anchoring of pitch movements: autosegmental association or gestural coordination? *Italian Journal of Linguistics* **18**. 19–38.
- Ladd, D. Robert (2008). *Intonational phonology*. 2nd edn. Cambridge: Cambridge University Press.
- Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner & Astrid Schepman (1999). Constant ‘segmental anchoring’ of F_0 movements under changes in speech rate. *JASA* **106**. 1543–1554.
- Ladd, D. Robert, Ineke Mennen & Astrid Schepman (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *JASA* **107**. 2685–2696.
- Marin, Stefania & Marianne Pouplier (2010). Temporal organization of complex onsets and codas in American English: testing the predictions of a gesture coupling model. *Motor Control* **14**. 380–407.
- Marin, Stefania & Marianne Pouplier (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *JPh* **42**. 24–36.
- Maslow, Abraham H. (1966). *The psychology of science: a reconnaissance*. South Bend: Gateway.
- Mücke, Doris, Martine Grice, Johannes Becker & Anne Hermes (2009). Sources of variation in tonal alignment: evidence from acoustic and kinematic data. *JPh* **37**. 321–338.
- Mücke, Doris, Jagoda Sieczkowska, Henrik Niemann, Martine Grice & Grzegorz Dogil (2010). Voicing profiles, gestural coordination and phonological licensing: obstruent-sonorant clusters in Polish. Poster presented at the 12th Conference on Laboratory Phonology, Albuquerque.
- Mücke, Doris, Martine Grice & Taehong Cho (2014). More than a magic moment: paving the way for dynamics of articulation and prosodic structure. *JPh* **44**. 1–7.
- Mücke, Doris, Anne Hermes & Taehong Cho (2017). Mechanisms of regulation in speech: linguistic structure and physical control system. *JPh* **64**. 1–7.
- Mücke, Doris, Anne Hermes, Timo B. Roettger, Johannes Becker, Henrik Niemann, Till A. Dembek, Lars Timmermann, Veerle Visser-Vandewalle, Gereon R. Fink, Martine Grice & Michael T. Barbe (2018). The effects of thalamic Deep Brain Stimulation on speech dynamics in patients with Essential Tremor: an articulo-graphic study. *PLoS One* **13**. <https://doi.org/10.1371/journal.pone.0191359>.
- Nam, Hosung, Louis Goldstein & Elliot Saltzman (2009). Self-organization of syllable structure: a coupled oscillator model. In François Pellegrino, Egidio Marisco, Ioana Chitoran & Christophe Coupé (eds.) *Approaches to phonological complexity*. Berlin & New York: Mouton de Gruyter. 299–328.
- Ohala, John J. (1990). There is no interface between phonetics and phonology: a personal view. *JPh* **18**. 153–171.
- Pastätter, Manfred & Marianne Pouplier (2015). Onset-vowel timing as a function of coarticulation resistance: evidence from articulatory data. In The Scottish Consortium for ICPhS 2015 (ed.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow. Available (January 2020) at <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0783.pdf>.
- Pierrehumbert, Janet B., Mary E. Beckman & D. Robert Ladd (2000). Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Philip

- Carr & Gerard Docherty (eds.) *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press. 273–303.
- Port, Robert F., Jonathan Dalby & Michael O'Dell (1987). Evidence for mora timing in Japanese. *JASA* **81**. 1574–1585.
- Poupier, Marianne (2012). The gestural approach to syllable structure: universal, language- and cluster-specific aspects. In Susanne Fuchs, Melanie Weirich, Daniel Pape & Pascal Perrier (eds.) *Speech planning and dynamics*. Frankfurt am Main: Lang. 63–96.
- Prieto, Pilar, Mariapaola D'Imperio & Barbara Gili Fivela (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech* **48**. 359–396.
- Prieto, Pilar & Francisco Torreira (2007). The segmental anchoring hypothesis revisited: syllable structure and speech rate effects on peak timing in Spanish. *JPh* **35**. 473–500.
- Ramus, Franck, Marina Nespor & Jacques Mehler (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* **73**. 265–292.
- Saltzman, Elliot & J. A. Scott Kelso (1987). Skilled actions: a task dynamic approach. *Psychological Review* **94**. 84–106.
- Saltzman, Elliot & Kevin G. Munhall (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* **1**. 333–382.
- Saltzman, Elliot, Hosung Nam, Jelena Krivokapić & Louis Goldstein (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the 4th Conference on Speech Prosody*. Campinas, Brazil. 175–184. Available (January 2020) at <http://www.isca-speech.org/archive/sp2008/>.
- Seidler, Rachael D., Jay L. Alberts & George E. Stelmach (2002). Changes in multi-joint performance with age. *Motor Control* **6**. 19–31.
- Shaw, Jason A. & Adamantios I. Gafos (2015). Stochastic time models of syllable structure. *PLoS One* **10**. <https://doi.org/10.1371/journal.pone.0124714>.
- Shaw, Jason A., Adamantios I. Gafos, Philip Hoole & Chakir Zeroual (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology* **26**. 187–215.
- Shaw, Jason A., Adamantios I. Gafos, Philip Hoole & Chakir Zeroual (2011). Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology* **28**. 455–490.
- Spivey, Michael (2007). *The continuity of mind*. Oxford: Oxford University Press.
- Tilsen, Sam (2016). Selection and coordination: the articulatory basis for the emergence of phonological structure. *JPh* **55**. 53–77.
- Tilsen, Sam (2017). Exertive modulation of speech and articulatory phasing. *JPh* **64**. 34–50.
- Tilsen, Sam (2018). Three mechanisms for modeling articulation: selection, coordination, and intention. *Cornell Working Papers in Phonetics and Phonology*. Available (January 2020) at <http://conf.ling.cornell.edu/~tilsen/papers/Tilsen%20-%20202018%20-%20Three%20mechanisms%20for%20modeling%20articulation.pdf>.
- Tilsen, Sam & Amalia Arvaniti (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *JASA* **134**. 628–639.
- Tilsen, Sam, Draga Zec, Christina Bjorndahl, Becky Butler, Marie-Josée L'Esperance, Alison Fisher, Linda Heimisdottir, Margaret Renwick & Chelsea Sanker (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology 2012*. 51–81.
- Turk, Alice & Stephanie Shattuck-Hufnagel (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology* **4**. 93–118.

Vatikiotis-Bateson, Eric, Adriano Vilela Barbosa & Catherine T. Best (2014). Articulatory coordination of two vocal tracts. *JPh* **44**. 167–181.

Waltl, Susanne & Stefania Marin (2010). Temporal organization of three-consonant onset clusters in American English. Poster presented at the 12th Conference on Laboratory Phonology, Albuquerque, New Mexico.

Wiese, Richard (1996). *The phonology of German*. Oxford: Clarendon.