ARTICLE

Check for updates

The mental representation of sounds in speech sound disorders

Soujanya Pathi [1✉] & Prakash Mondal [1]

The objective of this study is to investigate facets of the human phonological system in an attempt to elucidate the special nature of mental representations and operations underlying some of the errors in speech sound disorders (SSDs). After examining different theories on the mental representations of sounds and their organization in SSDs, we arrive at the conclusion that the existing elucidations on the phonological representations do not suffice to explain some distinctive facets of SSDs. Here, we endorse a hypothesis in favor of representationalism but offer an alternative conceptualization of the phonological representations (PR). We argue that the PR is to be understood in terms of a phonological base that holds information about a segment's acoustic structure, and which interacts with other levels in the speech sound system in the mind so as to produce a certain sound. We also propose that the PR is connected to an interface module which mediates interactions between the PR and the articulatory system (AS) responsible for the physical manifestation of speech sounds in real time by way of the coordination of activities of speech organs in the vocal tract. We specifically consider different stages of operations within the interface, a specialized system within the cognitive system, which can explain patterns in the SSD data that have so far remained elusive. Positioned between the PR and the AS, the interface module is the heart of the current study. The presence of an interface module is necessitated by the fact that not all errors of SSDs are explainable in terms of structural, motor or even the symbolic misrepresentations at the level of PR. The interface acts as a mediating system mapping sound representations onto articulatory instructions for the actual production of sounds. The interface module can receive, process, and share the phonological inputs with other levels within the speech sound system. We believe an interface module such as ours holds the key to explaining *at least* certain speech disarticulations in SSDs.

[1] Indian Institute of Technology Hyderabad, Hyderabad, India. ✉email: la17resch01002@iith.ac.in

Introduction

L anguage is fundamentally a complex system well capable of regulating and nurturing our social behavior, actions, and thoughts. This understanding has brought several significant questions and issues on the relationship between language and cognition to the forefront. This line of inquiry marks a significant contribution not only to linguistics and psychology but also to cognitive science that aims to develop all-encompassing theories of various human abilities, including language. From Chomsky's (1975) conception of language as akin to a mental organ to viewing language from evolutionary, neurobiological, and psychological perspectives, our understanding of language has come a long way. In understanding the theoretical foundations of the study of the human mind, the analysis of speech production has in fact played a crucial role in first formulating cognitive theories and then subsequently applying them in the real-world context. There are, however, certain gaps in our understanding of speech production, more specifically in speech sound disorders (SSDs), that elude satisfactory explanations. Though there is a significant amount of research claiming a strong relationship between the mental representations and SSDs, there is little evidence to show how exactly these representations hinder speech production. There have been several attempts, of course, at explaining how the weakness in quality and accessibility of phonological representations (PR) may eventually result in unintelligibility of speech (Anthony et al., 2011). This explanation, however, is limited in a manner that is applicable only to a subset of the whole range of speech dysfunctionalities that we have at our disposal. It does not also suffice to explain, for instance, cases where a child exhibits speech abnormalities despite having a non-defective PR. If PRs, as popularly advocated by many (Dodd, 2005; McNeill and Hesketh, 2010; Antony et al., 2011; Sutherland and Gillon, 2007) are to be involved in the misrepresentations of sounds, then it simply does not explain why and how a child can possibly possess the capacity to distinguish a minimal pair set (for example, the difference between 'pat' and 'bat'). Likewise, if we are to point to the inaccessibility of the mentally instantiated phonological symbol as the reason for speech dysfunctionalities, it does not always seem reasonable to attribute all the speech problems to a loss of certain cognitive capacity that cannot, for reasons that are not very clear, access the right symbol from the phonological system. Moreover, given the understanding that speech disorders mostly exhibit a pattern in terms of the errors they commit, the symbol extraction problem does not seem to explicate the exact reasons as to why only certain specific sounds (under predictable environments) malfunction. It is cases like these that we believe require closer scrutiny and probably more explanation in terms of what kind of cognitive processes drive a child to produce a certain sound in a way that is deviant from the typical speech. Therefore, with this paper, we intend to make forays into the cognitively instantiated speech sound system that can hopefully project a window into at least some speech problems. For the same, we have advanced a cognitive model of sound representations that focuses on understanding the internal processes that ultimately lead to variants in speech production in atypical populations with SSDs. The proposed cognitive model essentially consists of two levels: the PR and the interface. The articulatory system (AS) is not a part of the cognitive model as it only represents an output system towards the endpoint of the sensory-motor continuum. The AS, however, is connected to the interface and is crucial for the actual realization of sounds in a real-time world (see Fig. 1). While each of these levels plays a pivotal role in our cognitive model, the interface, sandwiched between the PR and the AS, shall specifically remain our primary concern. In the coming sections, we shall discuss in detail, with the help of relevant data, how a model such as ours can account for

variations in speech differences in atypical populations with SSDs.

The paper is structured as follows. Section "Speech sound disorders and cognition" presents a review of the literature on SSDs, and their position in our present study. Section "Phonological representations" begins by stating the definition of the PR as conceived of by earlier scholars and gradually navigates to a discussion of the PR as formulated in the present paper. The novelty of the PR is primarily presented in terms of the variant linguistic inputs the system processes, the manner in which certain operations are performed, and its interactions with the other levels in our cognitive model. Section "The cognitive model" presents a detailed description of the interface and the operations that take place here. Starting with the basic characterization of how an interface module is described in the current study, this section also offers, with the help of diagrammatic representations, a detailed view of the different levels and the various operations that occur within the interface. This section also highlights how a 'miscalculation' in linking the sounds to their relevant features at the level of an interface may eventually lead to dysfunctional speech. Section "The articulatory system" discusses how the AS works. This is followed by the results of applying the model to the relevant data from SSDs as furnished in section "Samples of data". Sections "Implications" and "Concluding remarks" discuss implications and conclusions, respectively.

SSDs and cognition

SSDs is an umbrella term referring to any kind of disarticulations pertaining to speech perception, production, and mental representations of sounds and other speech units. SSDs may be described as ranging from something as "mild" as a lisp (interdentalizing the /s/ sounds, sometimes identified via the substitution of a voiceless [θ] sound for an [s]) to a disorder as significant as that found in an individual who is completely unintelligible (Bernthal et al., 2017). While some diagnosis and intervention studies have proved effective in investigating the relationship between speech and other modes of expressions such as writing (Stackhouse et al., 2006), we primarily intend to investigate the relationship between speech sounds, as produced in atypical populations, and their mental representations. Nested within the cover term, SSDs are different classification terms such as articulatory disorders, phonological disorders, childhood apraxia of speech, etc. In this paper, however, we restrict ourselves to focusing on the disordered phonological systems that result in the production of deficient and unintelligible speech
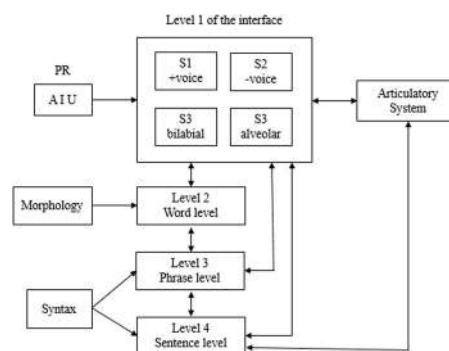

Fig. 1 The cognitive model: PR, the interface and the AS. The input fed by the PR is processed by the interface at different levels before forwarding it to the AS. Morphology and syntax are shown connected to the interface to indicate any word-based or phrase-based modifications that may indirectly feed into the interface.

utterances. That is, SSDs impacting the way speech sounds or segments function within a language shall remain our prime focus. In this connection, we care to emphasize that speech errors occurring due to physical abnormalities are outside the purview of the present study. We aim to present only a theoretical explanation of the possibilities of errors that lie within the speech sound system in mind and relate them to SSDs, from a view, again, constrained within a segmental approach. The relationship between cognition and speech production has been a subject of many years of investigations. Unlike cases of physical deformity, the cognitive effects on speech production are less apparent, thereby making them more difficult to diagnose. The relation between cognition and speech production, however, is irrefutably evident. A vast amount of work on SSDs and underlying representations indicates that children with SSDs cannot produce speech correctly because they have poorly specified PR (McNeill and Hesketh, 2010; Antony et al., 2011; Sutherland and Gillon, 2007). That is, a poorly developed PR may impede a child's ability to discriminate between sounds that share similar articulatory features (which are usually called distinctive features (DFs)). Individuals vary in how exactly they code the phonological information of words and also in how readily they access PR of words (e.g., Anthony et al., 2009, 2011). Furthermore, there is a growing body of evidence indicating that the representation-related phonological abilities of children with SSDs tend to function poorly on both expressive and receptive measures of phonological representation (Edwards et al., 1999, 2002; Hoffman et al., 1985; Munson et al., 2005; Rvachew et al., 2003; Sutherland and Gillon, 2005, 2007). Hence, there is enough evidence to support the claim that PR and SSDs are related in some way—a claim which seemingly indicates that the psychological factors or certain facets of cognition play a crucial role in SSDs. A similar understanding was echoed in a case study conducted by Leahy and Dodd (1987). Bizarre phonological processes in a child, such as the deletion of final consonants or marking them by a glottal stop, the marking of intervocalic consonants by a glottal stop, the use of a non-English fricative to mark consonant clusters were traced back to the cognitive abilities. All of these problems are typically found in phonologically disordered children. The child in question, nonetheless, exhibited no form of physical abnormalities. In concordance with Leahy and Dodd's work, Shriberg and Widder's (1990) findings from nearly four decades of speech research in cognitive impairment indicated that persons with cognitive impairments or any sort of deficit at a mental level are likely to have speech problems. That is, the articulatory skills of a subgroup with cognitive deficits differed significantly from normal developing children. Similarly, in several of the case studies conducted by Sutherland (2006) on children with severe speech impairments, it has been observed that 3 out of 4 children demonstrated poor phonological skills. This indicated that children with deviant consistent speech impairment experienced deficits at the cognitive-linguistic level (i.e., phonological representation) of speech production (Dodd, 2005). Though all the cases acknowledge that cognitive impairment may eventually lead to problems in speech productions, it is not yet clear how cognition can lend itself to explaining specific speech impairments. There are some studies which have provided a more succinct view on PRs and how they relate to speech difficulties (McNeill and Hesketh, 2010; Antony et al., 2011; Sutherland and Gillon, 2007). These studies, however, fail to explain scenarios where the speech sound discriminatory ability works perfectly fine. Therefore, regarding PR as the only mental base may not be a viable option. To that end, we propose that the phonemic information is specified at different levels within the speech sound system of the cognitive system, and speech impairments could be triggered for various miscalculations that many happen at levels other than the

PR (for example, the interface system proposed here). The nature and the function of the interface are discussed in detail in the coming sections.

Phonological representations

Traditionally, the term 'phonological representations' refers to the underlying sound structure of specific words stored in the long-term memory (Locke, 1985). It refers to the abstract system of phonological knowledge, a representational domain, that aids word learning, speech production, and literacy development. A system of symbols and representations, PRs are assumed to become more nuanced with age. Walley et al. (2003) propose that children's PR is more holistic, underspecified, and therefore may lead to a less discriminatory skill among the vocabulary items. Allophones, for instance, refer to a specific realization of a certain sound (a phone). While the sound /p/ occurring in the initial syllable of a word is pronounced with an aspiration, the sound elsewhere is produced without an aspiration. This discriminatory nature of sounds may or may not be available in a child's PR. These nuances, however, are slowly and gradually acquired at the later stages. PR is considered vital within the group of children with SSDs because there is considerable evidence that the weakness in establishing and accessing the PRs may often lead to speech difficulties among children with SSDs (Anthony et al., 2011; Sutherland and Gillon, 2007). While the present study most certainly cements the belief that there is a correlation between the PR and the speech disorders, it does not, for reasons that are discussed later, agree with the traditionalist view that PR has all the necessary information pertaining to sounds. Conventionally, it is understood that the information that is present in the PR is described in terms of DFs that correspond to the articulatory features (Berent, 2013). DFs are binary features reflecting articulatory properties or aspects of sounds. Features such as ±voice, ±coronal, ±round, ±back, etc. provide the information on specific sounds (Chomsky and Halle, 1968). The very premise of the present theorization is based on the postulate of multiple levels of specification of sound representations and also on the idea that PR is mentally encoded as acoustic traces (Coleman, 2002). It is not plausible to view PR as a container that holds all the phonetic details of a segment. Rather, the present study adopts the concept of an element in accordance with the Element Theory or ET (Kaye and Harris, 1990; Harris, 1994; Backley, 2011), which was developed as part of Government Phonology (Kaye et al., 1990). The adoption of the ET for the current framework is motivated by the ET's conception of sound elements positing a non-traditionalist, non-articulatory view of linguistic sounds. ET essentially differs from the earlier or the more popular works on DFs in terms of what they consider to be the building blocks of the phonological structure. ET supposes that the 'elements', each of which corresponds to an ensemble of acoustic properties, are the building blocks of a phonological structure. This assumption marks a stark contrast from the view of DFs, wherein the features are essentially linked to articulatory phonetics. Since articulatory features and properties are sort of *reified* in DFs, the acoustic view of sound elements in Element Theory offers the desired advantage we need for the cognitive model proposed. Our emphasis on the appropriateness of the ET for the present cognitive model also centers around this major difference. The cognitive architecture of the model is such that the segmental data is distributed at various levels. Acoustic properties of the elements at PR ensure that no articulatory gestures occur at this stage. An assumption such as this also accounts for those errors in speech disorders which appear despite the PR remaining more or less intact. The test of this can be conducted by a task on phonetic discrimination in which children end up making speech errors

despite good PR skills. Therefore, we claim that ET seemingly offers a slightly more efficacious solution to certain of the speech errors, which cannot be simply explained by the symbolic mis-representations at the level of the PR.

To elaborate on the notion of PR, we point out that it can be viewed as a system that contains certain elements, the combination of which gives rise to more complex segments. It consists of a set of finite elements like |A|,|I|,|U| that correspond to different acoustic properties. The element |A|, for instance, typically correlates to central spectral energy mass where high F1 converges with F2. Similarly, |I| corresponds to a high spectral peak where high F2 converges with F3, and |U| corresponds to a low spectral peak where low F2 converges with F1. The phonetic realizations of these elements can be perceived as non-high, front and rounded vowels, respectively, when they occur in a nucleus position. Likewise, when consonants occur in an onset position, the phonetic realization of |A| can be manifested in and as pharyngeal, uvular, and laminal coronal consonants; the phonetic realization of element |I| as palatal and apical coronal consonants, and the phonetic realization of element |U| as labial and velar consonants. Thus, it is possible to encode sound segments in terms of these elements and/or combinations thereof. The sound /æ/, for example, can be viewed as a combination of the elements |A| and |I| as it represents lowness (non-high) and frontedness. In the same way, the sound /ɒ/ can be viewed as a combination of the elements |A| and |U| as the sound represents both lowness (non-high) and roundedness. Now, for the purpose of understanding how these elements function at a later stage, let us take any two elements, say |X| and |Y| (where |X| and |Y| represent any elements in the PR) that combine to give rise to a segment, say, /z/. The resultant segment /z/, formed by the combined acoustic properties of two or more elements, is fed into a system where it takes another form. Although the elements in PR tell the speakers which patterns they must produce, it does not tell the speaker how to produce them. The description of the sounds in terms of the movements of the articulators in the vocal tract and the ways in which these constrain the production of speech sounds physically are all realized at a stage beyond the PR. For now, the function of the PR is to simply provide the under-specified inputs (the segment /z/) to the system where the segments can be further processed.

What contributes to a deficiency in speaking and reading is often attributed to primarily having deficits at the level of the PR. Most of the research claims that children with speech impairments produce erroneous segments or sounds because their PR, by default, is disrupted. That is to say that the PR of a person with a speech disability and that of a person without a disability are different. In other words, it seems to suggest that the disabled have a defective PR and the abled a perfect PR. But more often than not, we see that phonological phenomena such as metathesis and spoonerisms are not uncommon in persons without speech disorders. While one can argue that they are mere 'slips-of-the-tongue' and therefore occur due to the articulatory factors rather than to the representational factors, it is also worth noting that these 'slips-of-the-tongue' also often provide useful insights into the phonological structure of language (Harrikari, 1999). It is therefore suggested that it is not a viable option to directly dismiss or establish PR as being either imperfect in the case of SSDs, or totally perfect as in the case of persons without speech impairments. Rather, what seems to be more likely and also plausible is considering PR to be a representational system that is just 'good enough' in both cases. The correct or incorrect utterances produced at the level of the AS are not because of 'mental mis-representations' at the PR level but have a different origin.

PR plays a special role in the cognitive system as a whole by virtue of its functionality and the prominence it holds in our model. As mentioned earlier, PR, as viewed in this current paper, differs considerably from the earlier conceptions of PR in terms of its content and functions. PR here is not merely viewed as a repository of symbolic expressions but as a processing system with a set of operations. Though the contents of the PR in our view vary from those in the current literature, the contents of PR are just like other sensory representations, but the tasks for discerning the nature of PR are, of course, to be understood in terms of perception and discrimination at an acoustic level. The tasks designed to check the nature of PR are indeed perceptual in nature. This suggests that the contents of PR conceived of in acoustic terms in the present paper can be organized in terms of contours of perceptual maps of sounds. But, since linguistic sounds need to be produced as well, the elements of PR in speech production tend to align well with contours of perceptual maps of sounds, somewhat along the line of thinking in the analysis-by-synthesis model (Stevens and Hanson, 2010). This helps unite not only action and perception but also perception and cognition (Idsardi, 2015). The dichotomy between cognition and perception dissolves when one notes that cognitive representations often have a pervasive influence over perceptions in all sorts of perceptual processing (including, of course, speech processing) while being organized by perceptual maps in a paradigm of predictive coding (Clark, 2013, 2016).

The cognitive model

The interface. Dissociations of the orthographic input from the speech output presuppose distinct processes and mechanisms. Attempts to recognize and understand the information processing system used for reading, writing, and oral spelling led to investigations, particularly into the systems involved in transforming the text to speech. The dual-route conception of reading, which gained popularity in the 1970s, was also an attempt to understand the pathways that transform orthographic representations into sounds. The conception was also further advanced to deal with various questions relating to different profiles of deficits such as surface dyslexia. (Ellis and Young, 1996). In view of these developments, Ellis and Young explore the possibility of the deficit having arisen at the orthographic level, rather than at any other level, which most likely affects the whole word reading mechanism. In general, the model assumes that the relevant processes transform an orthographic code into a phonological code and subsequently lead to 'speech'. It was also revealed by Humphreys and Evett (1985) that 'speech' occurs as a consequence of the grapheme–phoneme conversion. The dual-route conception of speech sound production can also be found to be elaborated on in Levelt's (1989) model of language production. It is, however, not clear how the impediments during the course of speech actualization or during the physical manifestation of the individual sound segments are dealt with in the dual-route model because it assumes coarse-grained representations of sounds. In contrast, the present model is much more granular in dealing with the decomposition of the segments themselves. The present conception of an interface is also an attempt to decipher the inner workings of the speech sound system and thereby determine the possible transformations that reflect speech impairments in SSDs. The misrepresentations can come about not only in terms of a linear progression of a sequence of steps involved in the production of speech sounds, but also in terms of the possible mis-mappings it can face in the production of sounds. Apart from that, one may also wonder if an interface system is actually about phoneme-to-motor routine conversions. The answer is partly 'yes' because the interface is plausibly located in the phonological network responsible for the mapping of sound representations onto articulatory instructions to be ultimately routed through the

motor cortex to the articulators. However, the interface in the present model is not supposed to deal with the neural operations directly located in the motor cortex. Significantly, given that in the dual-route model and also in Levelt's model, the segmental and metrical information is provided in a readymade manner in the phonological output lexicon, the interface system in the current model perhaps functions at a level *prior to* the operations in the phonological output lexicon due to the granular representation of sounds couched in terms of elements.

Similar considerations also apply to the speech processing model of Stackhouse and Well (1997) and related processing models (see Dodd, 2005). For example, Stackhouse and Well's speech processing model has two connected sub-systems—one for phonological representations (that is, the PR) and another for motor programs further linked to the nodes of motor programming (for the creation of motor programs of new words) and planning (for the motor outputs of sounds). This is indeed a significant processing linkage because this can explain how children with speech-language difficulties can also have poor skills on phonological awareness (especially on segment-level tasks) (Schaefer et al., 2016). This is perhaps mediated by mapping issues at the link between phonological recognition and the PR. This consideration notwithstanding, the PR in this model is coarse-grained, and no further level between the PR and motor programs is presupposed. Hence it is not clear how (abstract) sound representations can be converted into motor programs. Even though motor programs can be thought of as motor schemas, these schemas have to be mapped onto the right combination of sound properties. Since most speech errors (in SSDs) consist of alterations in specific acoustic and articulatory features, it is unclear how motor schemas can target such specific sound features. Moreover, the interface system in its functioning significantly differs from Dell's connectionist model of spreading activations (Dell, 1986) because, in Dell's model, there is no scope for the mapping of the symbolic units of phonology onto articulatory instructions. Besides, phonological units are not decomposed into their components—acoustic or otherwise, although the significance of phonemes is acknowledged (Dell, 1986). Hence, a cognitive model such as ours offers insights into the connections between the phonological system and SSDs.

The concept of the interface, which forms the locus of the current model, is defined as a transducing system that receives the inputs from the PR and processes/alters the input obtained from the PR before finally sharing it with the AS. This work chooses to use the term 'interface' to designate a system in itself in line with Jackendoff's (2002) concept of interface systems. The modular view of grammar and the interface processes have received much attention in recent times, partly motivated by Fodor's (1983) conception of modularity. It is based on the premise that distinct types of grammatical processes impose different rules autonomously, largely independently of the function of the other modules. A word order generalization from a syntactic domain, for instance, may not have its counterpart in the semantic or phonological system. So the content in each of the domains of language is unique (a matter of *domain specificity*). Thus, it seems natural to assume that these modules are autonomous and function separately. But given that all linguistic expressions uttered are products of not one single module but several modules like morphology, phonetics, syntax, etc., it is likely that all these modules work together to assemble linguistic structures, including sound structures. There is a significant amount of work in phonology that revolves around the issue of exploring the extent to which phonological processes are guided by articulatory and perceptual (i.e. phonetic) considerations (Ohala, 1974, 1983; Archangeli and Pulleybank, 1994; Steriade, 1995; Jun, 1995; Kaun, 1995; Flemming, 1995; Silverman, 1995; Kirchner, 1998).

Similarly, phonological processes are sensitive to morphological structures as well. The interactions between prosodic, syntactic, semantic, and pragmatic systems have also invited focus in the current research (Ramchand and Reiss, 2007). Accordingly, one can also distinguish different types of interfaces like the syntax–semantics interface, the syntax–phonology interface, etc.

The notion of the interface is primarily motivated by the need to look at points of cross-talk between modules where the idea of autonomy somewhat dissolves. In partial agreement with the works cited above, we draw upon the notion of an interface as a system that mediates not between two different domains of language but between the PR and the AS within a narrow spectrum of the speech sound system. Therefore, we suggest that the mentally instantiated phonological system consists of some well-designated components, one of which is the interface. The notion of an interface has also been discussed in detail by Reiss and Volenec (2018) with reference to the mapping between phonology and phonetics. The proposal in Reiss and Volenec is considerably new because it does not simply espouse DFs with reified articulatory features, as is found in generative phonology. The paper also talks about the transduced features, such as $PR_{[ROUND]}$ or $PR_{[+BACK]}$ in terms of the muscular contractions that each of them relates to. It is, however, still unclear what exactly triggers the articulatory movements for the specific sounds. Reiss and Volenec, of course, discuss how these temporally coordinated muscles are related to features, but it is not clear by what means the information specified in abstract DFs gets translated into the actual rounding of the lips (in the case of $PR_{[ROUND]}$) or the real-time function of raising the back of the tongue (in the case of $PR_{[+BACK]}$). In other words, what articulatory aspects get encoded in each specific feature are not clearly established, and also how these features interact with specific muscles for articulation. But for the purpose of the externalization of sound representations, there has to be a mechanism that explicitly states or provides at least some kind of a signal for the articulatory movements to get started. In trying to address this key issue, the current hypothesis has adopted a view of an interface that not only bridges the gap between the abstract mental representations of sounds and the actualizations of the sounds but also states with the help of a series of steps what exact instructions have to be followed in order to produce a particular sound.

The interface system is *informationally encapsulated* (Fodor, 1983) relative to syntax and semantics, for example, because only phonological objects and the internal grammar for operations on such objects are relied on. However, the interface system is not *informationally encapsulated* relative to the sub-systems of the entire speech sound system because it interacts with the system of PR, for no sound can be produced in isolation. It is for this very reason that we contend that the interface module interacts with some specialized sub-systems responsible for determining whether the outputs of the interface can be saliently affected by morphological, phrasal, and syntactic rules of formulation in connected speech. This is so because sound alterations in connected speech can be sometimes modulated by language-specific morphological and syntactic rules. Hence we distinguish one level within the speech sound system from another. The term 'levels' here indicates specialized sub-systems that execute different sets of specific operations and rules. In this paper, 'levels' are used to denote not just the interface sandwiched between the PR and the AS but also each of sub-systems (levels 2, 3, 4, etc.) that help determine the final output to be sent to the AS through word-level, phrase-level, and sentence-level modifications. In this sense, the core outputs of the interface module are accessible to morphological and syntactic rules so that specific sound alterations can be (sometimes) morphologically and

syntactically motivated. This by no means suggests that all of morphology or syntax is relevant to the operations of the interface. Although morphological and syntactic rules do not directly affect the operations of the interface, to be discussed below, they may indirectly feed information about word-based or phrase-based modifications to sounds into the interface. It is thus eminently compatible with Jackendoff's (2002) conception of *representational modularity* that relaxes Fodor's criterion of information capsulation. So the other levels of the interface system (levels 2, 3, 4, etc.) are merely sub-systems of mapping of those pieces of structure that relate some aspects of morphology and syntax to certain aspects of sounds. In any event, we wish to make it clear that the study of the interaction between the interface and the other domains such as syntax, morphology, etc., is outside the purview of the present study.

Decoding at level 1 of the interface in typical populations. As mentioned already, the interface is a complex system capable of performing several operations. Level 1 of the interface system as a whole, however, acts as the core domain where the primary operations are executed. As was mentioned in section "Phonological representations", PR comprises a set of finite elements like /A/, /I/, /U/, each of which encodes different acoustic properties. These elements do not encode the articulatory properties of the sounds, and hence, PR does not provide any articulatory information about sound segments. This crucial piece of information is provided at the next level, i.e., at the interface. If we are to comprehend how the interface functions, it is imperative to understand the several operations that take place within level 1 of the cognitive model in detail. In the process, we will clarify how the linguistic input from the PR is shared with the interface.

The very first level of the interface system as a whole (see Fig. 1) hosts a set of slots comprising articulatory features or AFs (which were traditionally perceived as DFs) pertaining to both consonants and vowels. The preference for the term AFs over the traditionally more prevalent term 'distinctive features' is mainly due to the subtle variations in the roles each of these exhibits. While both encompass fundamental properties of speech sounds, AFs in our cognitive model are more concrete and less abstract by being associated with specific articulatory instructions. AFs are supposed, in addition to providing a necessary basis for understanding the properties of sounds, to pass the articulatory instructions to other levels.

For each person, depending on the phonology of their language, the slots are allotted relevant articulatory features. For instance, the absence of a particular feature, say, a *bilabial* feature in the phonological system of that particular language entails that the slot reserved for the *bilabial* feature is transiently inactive. The slot is rendered only inactive and not completely absent from level 1 as one can always learn to produce a *bilabial* sound even when it is not present in their language. Thus, level 1 is said to comprise a finite number of slots, whose activeness or inactivity depends on the exposure to particular sounds. These slots can be thought of as motor schemas which, upon being linked to elements in the PR, activate instructions for the articulators to follow, somewhat along the line of thinking in articulatory phonology (Browman and Goldstein, 1989; Gafos, 2002). For example, if the intended sound is /s/ and the segment /x/, which is a combination of the elements /A/ and /H/, is rightly mapped on to slot 1 consisting of the *-voice* feature, it then results in the production of sound /s/. While the input received from the PR is fed to level 1 of the interface system, we suggest that it is usually the mapping or rather the *mis-mapping* of the input segment from the PR to the relevant slots that produce a defective speech.

Also, the connection between the PR and the slots is consolidated by certain aerodynamic factors. The gestural movements in different combinations and the aerodynamic factors not only influence the

formation of a particular sound but also explain the human tendency to prefer specific sound patterns over others. For instance, Ohala (1983) argues for the absence of the /g/ sound over the other voiced sounds in terms of the aerodynamic factors. He maintains that the sound /g/ is more susceptible to deletion than any other voiced plosive due to the degree of its closeness between the larynx and its point of closure. Because the location of the closure is much closer to the larynx, the air pressure in the supraglottal region exceeds that of the air below the larynx, thereby leaving insufficient air to drive the vibration in the vocal cords. Similarly, we suspect that a predisposition to using certain sounds over others can be traced back to the (mis)mapping of the input elements from the PR, which can be aerodynamically motivated. While most of the errors analyzed in the model are errors due to the mapping problem as part of the interface, these errors are not random, and hence they can arise as epiphenomena or 'side effects' of the processes within the AS. Because PR is solely a representative module devoid of articulatory slots, we presume that the articulatory or phonetic factors are *partly* instantiated in the interface (in slots) and *wholly* manifested in the AS. Consequently, a mis-mapping of the underspecified input from the PR onto the wrong slot evinces the instantiation of an unintended property of a specific sound, thereby resulting in the production of a disordered utterance. This kind of analysis is particularly helpful in analyzing the sound patterns in persons with SSDs since most of the errors produced in SSDs, if not all, dovetail with patterns indicating a preference for one sound or one class of sounds over others.

Decoding at level 1 of the interface in atypical populations. In the case of children with speech dysfunctionalities, it can be inferred that either they have a problem with keeping certain slots active, or there is a mismatch between the segments and the slots to which they are linked. To add further, it is also plausible that the segments shift to other unpredictable slots because the slots in which the segments were originally intended to fit were unavailable for various reasons. To illustrate this, let us have a look at the following example for the production of sound /f/ in the typical population. A diagrammatic representation of the same is presented in Fig. 2.

Step 1: U + H (from the PR) → /x/ where x is a resultant of the combination of two elements, carries some acoustic properties on its own but is devoid of articulatory properties.

Step 2: /x/ + S1 + S2 + S3 → /f/ where /f/ is the intended utterance.

For the atypical population (see Fig. 3),

Step 1: U + H (from the PR) → x where x is a resultant of the combination of two elements, carries some acoustic property on its own but is devoid of articulatory properties.

Step 2: /x/ + Sn → /z/ where n refers to some number other than 1 and z refers to an unintended utterance.
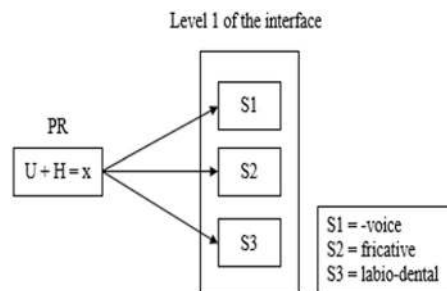

Fig. 2 Level 1 of the interface in typical populations. The figure indicates the correct mapping of the segment generated from the PR onto the right slots designated for the production of /f/.
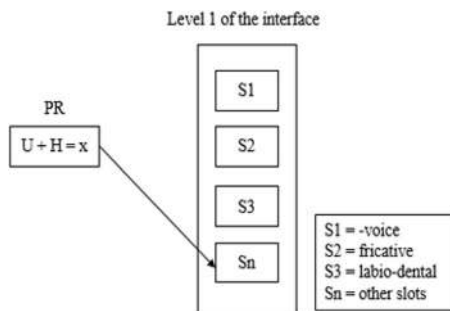
Level 1 of the interface

Fig. 3 Level 1 of the interface in atypical populations. The figure indicates the incorrect mapping of the segment generated from the PR onto the slots designated for the production of the sounds other than /f/.

Decoding at level 2, level 3 and level 4 connected to the interface module. The correct or incorrect utterance obtained from level 1 of the interface is put through several other levels, all connected to the interface in a bidirectional manner. The sound generated at level 1 is passed on to level 2 (word level), which checks for the neighboring sounds. If there is a chance of the segment getting altered, as in the case of co-articulation, it reverts to level 1 (since we mention that it is a bidirectional system) and picks up the required slot. The newly generated sound is again sent to the word level and further on. If the neighboring sounds do not affect the sound in any way, then they simply get carried on to the subsequent levels. The same is followed for level 3 (phrase level) and level 4 (sentence level). If the particular sound within the phrase or a sentence exerts an influence on the sound received from its previous levels, it is modified and then sent to the later stages. The need for such a multi-level module is necessitated by the fact that all the levels have different requirements to be fulfilled. Level 1, for instance, acts at a segmental level and, therefore, need not alter the specifications of sounds according to its environment. The question of an environment at the segmental level does not arise. Since only one segment is processed at a time, there is no possibility of another sound exerting an influence on it. But the case is not simple at other levels. Therefore, in order to cater to the different needs of the speech units, the interface is connected to other levels.

The set of operations that occur at different stations of the interface can be summarized as follows:

Step 1: Generate an underspecified segment from the PR; let us call it x.

Step 2: Pass x through level 1 of the interface

x gets itself attached to a slot Sn, where n is any number

$x + Sn = y$, where y is the sound with the articulation specified

Step 3: Pass y though level 2

IF no changes detected, THEN pass through level 3

IF changes detected, THEN revert to level 1

Select a new slot

Generate required sound, pass through level 2

Step 4: Pass through level 3

IF no changes detected, THEN pass to level 4

IF changes detected, THEN pass through level 1

Select a new slot

Generate required sound, pass through level 3

Step 5: Pass through level 4

IF no changes detected, THEN pass to the AS

IF changes detected, THEN pass to level 1

Select a new slot

Generate required sound, pass to AS

So far, the model has attempted to explain how sound shifts occur in the context of words, phrases, and sentences. But these shifts of sounds are not typical of SSDs as they frequently occur in

a typical population as well. However, one possibility that probably distinguishes the typical from the atypical population is the way in which these interactions occur. It is plausible that in atypical populations, the interaction between the levels occur even when the interaction is not required. That is to say that even in cases when there is no effect of the neighboring sound(s) at level 2, or even when there is no requirement for the generated segment to revert to level 1, it does so. We leave this matter open, though.

For the typical population,

IF $f \rightarrow f$, THEN move to level 3

For the disordered,

IF $f \rightarrow f$, THEN move to level 1

Since this is a generic model that accounts for both typical and atypical populations, there is a need to distinguish how the speech errors are actualized or manifested in the case of slips-of-tongue in typical populations from the way errors occur in disordered speech. Even though slips-of-tongue are usually associated with motor functions rather than with cognitive functions, it has also been stated in the recent research that they also get influenced by psycholinguistic mechanisms (Fromkin, 1971, 2012). One plausible explanation could be that in the case of disordered speech, the unnecessary interaction that takes place is, by default, permanent or more entrenched. For instance, if a sound /f/ is constantly mispronounced as sound /v/ in specific contexts, then it can be speculated that the mechanism that allows for the selection of the unintended slot at level 1, after the relevant segment comes back from level 2, is consolidated. Therefore, it can be said that it is the unnecessary interaction that is permanent in disordered speech and triggers the unintended utterance. In the case of slips-of-tongue, while the mechanisms that produce the unintended utterance remain the same, the mechanisms per se are not structured in a way that they are permanent. One of the major differences between slips-of-tongue and disordered speech is that the former does not occur frequently, but the latter does. Since phonological errors are not random errors and occur in patterns, it can be said that the mechanism or the interaction between the two levels (level 1 and level 2) itself is permanently altered.

The articulatory system

Though AS is not a part of the cognitive model per se, the system plays a crucial role in transforming the mentally instantiated segments into the sounds that can be actually realized and perceived in the physical world. This system is responsible for bringing together different speech articulators to produce the intended utterance by converting motor schemas from slots into articulatory instructions. This phase can roughly be thought of as the implementation level of Marr's (1982) three-level schema of the cognitive architecture. Both AS and Marr's level of implementation in a way deal with the physical realization of a representation. More specifically, in the case of the AS the abstract entities transduced from the interface are given a form that can be perceived by any human without any hearing impairments. For instance, if /p/ is the intended utterance, the AS is responsible for manipulating the speech organs such as the vocal folds, the upper lip, and the lower lip. In other words, the role of AS is to simply obtain the appropriate instructions from the interface module for a specific sound, coordinate the corresponding speech organs, and finally put them to use.

Samples of data

The secondary data collected from various sources which pertain to SSDs and also to the mental representation of sounds will be used below to illustrate the operations at the level of the interface

Table 1 Joseph's data.

Number	Target	Utterance	Sl. no.	Target	Utterance
1	Bite	bɑɪ	9	Kids	kɪp
2	Bus	bʌ	10	Mud	mʌ
3	Cheese	tɪ	11	Tooth	tʊʔ
4	Cut	kʌʔ	12	Drive	gɑɪ
5	Five	pɑɪ	13	Sharp	tɑʊp
6	Gift	gɪp	14	Soap	to
7	Toes	To	15	Sunny	tʌnɪ
8	Juice	dʊ	16	Soup	tʊʔ

Table 2 Josie's data.

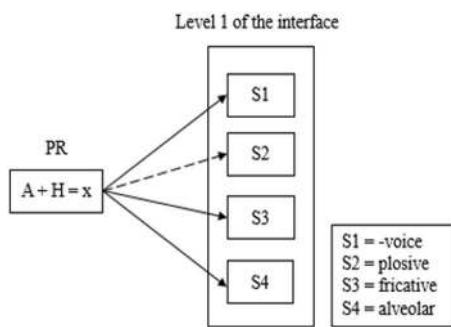| Number	Target	Utterance	Sl. no.	Target	Utterance
1	Cup	kʰʌ	11	Snake	fneɪʔ
2	Gone	kʰɒn	12	House	hæʊ
3	Knife	nɑɪ	13	Toe	tʰoʊ
4	Sharp	wja:	14	Mouth	maʊ
5	Fish	De	15	Nose	noʊ

Fig. 4 Substitution of /s/ with /t/ at level 1 of the interface. While the straight arrows represent correct mapping for the production of the sound /s/, the dotted arrows represent mapping of the incorrect slots.

and to determine the relevance the model bears on problems with speech production. The results of applying the proposed model to certain types of sound alterations in SSDs are described with implications for the cognitive representation of speech sounds.

Case 1. Presented in Table 1 is the clinical case conducted by Barlow and Gierut (2002) on a 4-year-old child Joseph diagnosed with functional phonological delay. The child displayed a variety of speech errors, a few of which are currently drawn from the large-scale study to illustrate how our proposed theoretical model can accommodate the actual data. The child in question displayed normal hearing, intelligence, oral-motor functioning, and regular receptive and expressive language skills as per the formal testing procedures. Joseph's speech data display several gaps in the normalized phonetic inventory of the English language and some deviant patterns that are otherwise not found. The kind of errors ranged from simple substitutions and deletions to cluster simplifications or a combination of all of these. We shall take one illustrative case. Figure 4 demonstrates level 1 of the interface in terms of how the substituted /t/ (/tʌnɪ/) for /s/ (sunny) can be produced, followed by their set of operations.

Operations
For the intended utterance /s/,
Step 1: A + H = x, where x is the underspecified segment from PR
Step 2: x + S1 + S3 + S4 = initiation of /s/ sound
For the disordered utterance /t/,
Step 1: A + H = x, where x is the underspecified segment from PR
Step 2: x + S1 + S2 + S4 = initiation of /t/ sound
While the sounds /s/ and /t/ differ minimally on a single slot, they also share the same place of articulation and voice, and yet the mis-mapping may result in a collapse of contrast between two sounds. As was also seen in Joseph's case, the sound /s/ never occurred in his phonemic inventory of sounds. Hence we can possibly infer that the mapping, or rather the mis-mapping of the

S2 slot from the underspecified PR segment, by way of fossilization, has permanently been established. The presence of the articulatory features and the mishaps in the operations performed at the level of the interface also serves as an explanation as to why Joseph's receptive skills are still intact (suggesting that the PR is good enough), despite his inability to produce the sounds correctly. Because the present model considers PR to be an efficient system with almost no malfunctions within it, we assume that Joseph still displays the capacity to understand /s/ and /t/ as two distinct sounds.

Case 2. In case 1, we have looked at errors of substitutions and their operations at the interface level. We will now look at how the model can explain the deletions. For that purpose, we will consider another set of sample data from a case study conducted on a subject named Josie between the ages of two and five (Bowen, 2015). Josie was diagnosed with developmental verbal dyspraxia (DVD) and had performed poorly on articulatory tests. Her speech was rendered unintelligible despite her maintaining a mid-range receptive, expressive, and total language score. The sound sequences for words used in Table 2 are impoverished and a part of the sample prior to the intervention.

Josie's disorder was severe, and often exhibited patterns that were most likely unintelligible. Though intervention studies altered Josie's speech at a later stage, for the purpose of our study, we shall first try to investigate what, in the first place, had caused such chronic distortions. Josie exhibited a range of patterns starting from single sound substitutions and deletion to the production of sounds that bore no resemblance to the target word. One possible explanation for the case of deleted sounds could be traced back to the inactivity in the slots. That is, there could be instances when the slots do not function actively, even in the cases when they are required to do so. The inactivity of a slot can eventually lead to two consequences: firstly, the segment generated from the PR, upon finding the slot inactive, deviates to other slots, thereby producing a different segment. So far, this mis-mapping has served as an explanation for the substituted sounds. Secondly, the segment generated from the PR, upon finding the slot inactive or invalid, does not end up being assigned any feature. However, in this case, the segment does not get itself 'attracted' to the wrong slot. Instead, the segment is left in situ, devoid of any articulatory features to process. Specifically concentrating on the case of /p/ deletion in the word 'cup', we speculate that the slots holding the corresponding features of /p/ fail to assign the articulatory features to the segment generated from the PR. Moreover, because the segment has been assigned a null value, no particular articulatory instruction is taken forward for the next levels. As a result of this, there is no production of the sound /p/ in the AS. The transitory nature of the slots also justifies why certain slots holding features like—voice remain passive in the production of /p/ but stay active in the production of other voiceless sounds like in the production of /f/. It is plausible that certain slots can go inactive for certain element combinations in this way due to the impact of relevant aerodynamic factors, as discussed in sub-sub-section "Decoding
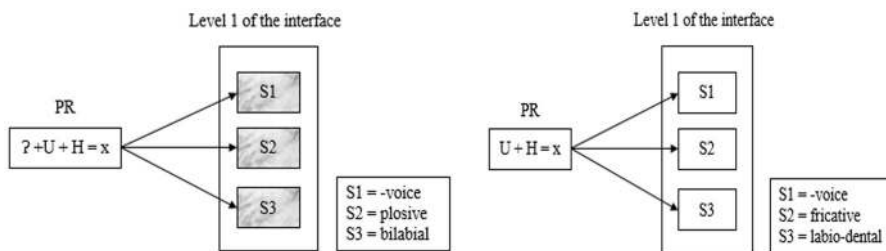
Fig. 5 a Inactive slots. The transiently inactive slots in the production of the sound /p/ are represented by the marbled slots. **b** Active slots. The slots for the production of the /f/ sound are represented by the plain slots.

Table 3 Data from Sutherland and Gillon (2005).				
Number	**Intended word**	**Uttered word**	**Word**	**Description**
1	tʃɜfou**t**	tʃɜfou**g**	'cherfoge'	Consonant change
2	kæt**ə**pɪlə	kæt**ʊ**pɪlə	'caterpillar'	Vowel change in an unstressed syllable
3	mæl**ə**tʃed	mæl**ʊ**tʃed	'melached	Vowel change
4	d**aɪ**nəsɒ	d**ʌ**nəsɒ	'dinosaur'	Vowel change in a stressed syllable

at level 1 of the interface in typical populations". Hence it is essentially the nature of the slots that give rise to the deletion and not the mapping. Illustrated in Fig. 5a and b are the inactive slots in /p/, and the active slots in /f/, respectively.

Case 3. Given below is some data pertaining to the articulatory defect. The data shown below corresponds to the numerous studies conducted on pre-school children with speech impairment. With the proposed model, we shall try and explain how the following speech inefficiencies can possibly occur.

One prominent feature of speech errors, which also makes it significantly different from the misarticulations produced by people with SSD's, is that they occur in patterns. Most of the speech errors, if not all, are bound to have some patterns. Data from Table 3 shows no signs of such patterns. The features of the sounds that are replaced barely match those of the intended sounds. Though there is an observed pattern in terms of which syllables (stressed or unstressed) or segments (either consonants or vowels) are generally prone to misarticulation in specific disabled individuals, it is difficult to identify otherwise what exactly prompts a shift in a sound which may be far removed from the originally intended sound. Barring a few features, none of the substituted vowels seem to indicate that they share a close resemblance to the originally intended sounds either in terms of height/backness of the vowel or in the rounding of the lips. Suppose we are to say that sound /t/ changes to /g/ as in the data confirmed by Sutherland and Gillon, the conventional or the traditional cognitive theories would render the sound change as being misrepresented completely in the PR. But then, if that were the case, the same sound change should occur everywhere regardless of the environment in which it is placed. Alternatively, if they were to argue that /t/ changes to /g/ because there is a significant difference in the voicing and the manner feature, it ultimately refers to the articulatory aspect of the sound and not to the mental representation as such. Such kind of speech disabilities are explained in terms of the interface module. Figure 6 shows the level 1 representation of what could be happening in the replacement of /t/ sound with /g/ sound.

Case 4. The case history of a 3-year-old child Kirk has revealed that the child had exhibited poor intelligibility in speaking despite having been identified with typical motor and language development. Hearing screening indicated that hearing was within
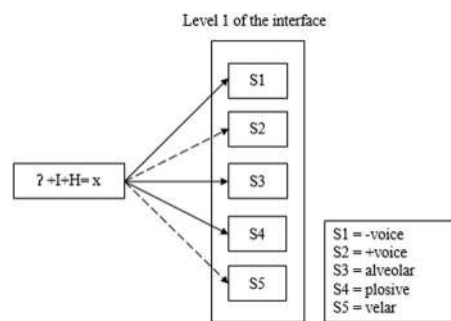

Fig. 6 Substitution of /t/ with /g/ sound. While the straight arrows represent the correct mapping for the production of the /t/ sound, the dotted arrows indicate the incorrect mapping.

Table 4 Kirk's data.					
Number	**Target**	**Utterance**	**Number**	**Target**	**Utterance**
1	p		12	z	d
2	b		13	n	
3	m		14	l	Ø
4	w	ø	15	ʃ	d
5	f	d	16	tʃ	d
6	v	d	17	dʒ	d
7	θ	d	18	j	Ø
8	ð	d	19	r	Ø
9	t		20	k	
10	d		21	g	
11	s	d	22	h	Ø

normal limits. A speech mechanism examination also indicated normal structure and function. Table 4 shows Kirk's transcriptions from single-word productions.

After careful scrutiny of the data, we arrive at certain inferences drawn from Kirk's data (Bernthal et al., 2017). This indicated speech inconsistencies of all types. Substitution errors occurred frequently and, more often than not, exhibited patterns indicative of preferences for certain sounds over others. Unusual processes like initial consonant deletions or final constant deletions, which
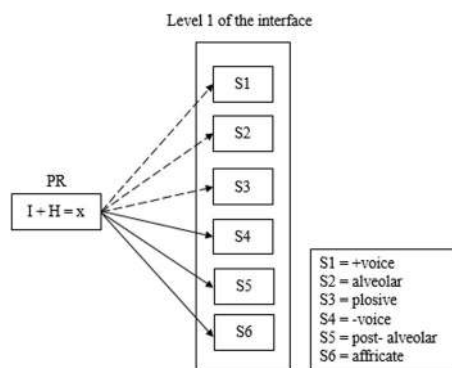
9

Fig. 7 Substitution of /ʧ/ with /d/ sound. While the straight arrows represent the correct mapping for the production of the /ʧ/ sound, the dotted arrows indicate the incorrect mapping.

are atypical for a 3-year-old child, were also observed. As far as the substitution errors are concerned, the analysis revealed that stopping was the most dominant and preferred processes of all. With /d/ substituting the likes of /f/, /v/, /θ/, /s/, /z/, /ʃ/, /tʃ/, and /dʒ/, the sound emerged as the most prominent one in Kirk's vocabulary. A similar pattern was also observed in Joseph's case where the child had also exhibited a similar preference for using plosives for fricatives and affricates (Fig. 7).

Operations
For the intended utterance /ʧ/
Step 1: I + H = x, where x is the underspecified segment
Step 2: x + S4 + S5 + S6 = initiation of /ʧ/ sound
For the disordered utterance /d/ (because all of the sounds are replaced by the sound /d/, it is safe to assume that Kirk produces the sound appropriately)
Step 1: I + N = x
Step 2: x + S1 + S2 + S3 = initiation of /d/ sound

One important alternative here is to appeal to articulatory gestures to explain the patterns of data above, as Namasivayam et al. (2020) have done. An articulatory phonology perspective on SSDs assumes that gesture hiding for homorganic gestures (involving common articulatory organs) and sometimes heterorganic gestures (involving distinct articulatory organs) may produce speech errors in SSDs. For instance, the individual sound pairs /g/ and /d/, or /w/ and /ʃ/ are supported by heterorganic gestures. But the problem is that the substitution of /g/ for /d/ in Joseph's data, or that of /w/ for /ʃ/ in Josie's data is not explained by heterorganic gestures. Namasivayam et al. when talking about specific SSD errors such as gliding and vocalization of liquids, stopping of fricatives, and also cluster reduction, have resorted to two kinds of explanation: gesture simplification and gesture overlaps for heterorganic gestures. Gesture simplification has been advanced for the gliding (and vocalization) of liquids, while gesture overlap has been advanced for heterorganic sounds in a consonant cluster wherein. However, it is easy to invoke gesture overlap because of the adjacency of the sounds supported by heterorganic gestures. The challenge the data in the current paper pose here for the gesture overlap explanation is that the sounds /g/ and /d/ in the case of the substitution of /g/ for /d/ in Joseph's data cannot be in an overlap in any sense (there is no velar sound in 'drive'). Likewise, the sounds /w/ and /ʃ/ in the substitution of /w/ for /ʃ/ in Josie's data cannot be said to be in an overlap (there is no velar sound in 'sharp', although the /r/ sound may have a slight labial component shared with the /w/ sound). The only remaining explanation is then gesture simplification for these substitutions (/ʃ/–> /w/ due to the difficulty in making the tongue-alveolar ridge constriction for fricatives, and /d/–>/g/ due to the difficulty in making the tongue blade constriction for

plosives). But it may be noted that an aerodynamic explanation for exactly such simplification may be supplied along the line outlined towards the end of sub-sub-section "Decoding at level 1 of the interface in typical populations".

Implications
The representative errors in SSD complexities, specifically those concerning the sound structure, stem from either 'misrepresented symbols' or from various processing deficits. Therefore, in order to be able to have a differential diagnosis and treatment therapies for the SSDs, the SSD classification must be efficiently established. Based on the earlier developments and the nuanced advances in neurolinguistics in later stages, several systems of SSD classifications have been proposed, some of which have had implications for the differential diagnosis and treatment planning. (Stackhouse and Wells, 1997; Waring and Knight, 2013; Shriberg et al., 2010; Dodd, 2014). However, these classifications, as Terband et al. (2019) claim, do not thoroughly explore the relationships between the different levels of causation, and hence, may deter efficient diagnosis, customize intervention, and optimize outcomes. The present cognitive model seeks to explore different levels in the phonological system and thereby identify the 'cause' of the speech deficit. The implications of such a model can primarily be derived from its core capabilities to recognize subtypes in the SSDs that qualify for a perfect PR and yet a defective speech output. The ability of SSD patients to identify and discriminate phonemes in relation to their ability to produce sounds, measured on standard clinical diagnostic tests, for instance, serves as a good predictor of the PR efficiency. Furthermore, the experimental validation of the model proposed, which is beyond the scope of the present paper, can provide a firmer ground for this.

Concluding remarks
The present study has attempted to demonstrate and explain why certain clinically notable segmental speech errors occur in people with SSD's that cannot be explained by significant sensory-motor impairments or impairments in mental representations. In a bid to explain why segmental errors occur, the current study has proposed a model of the interface wherein different stages of coding take place. It is suggested here that the interface module in itself comprises different levels wherein both simple and complex operations take place. It is also hypothesized that a possible miscalculation at any level of coding is what prompts an inaccurate or an unintended utterance. As of now, we are not currently certain if this particular model suffices for all kinds of segmental errors, but a further inquiry into a diverse range of experimental data can not only project interesting insights into the speech sound system within the cognitive system but also help in fine-tuning the present model.

Like every study, this particular study also has its own set of limitations. Firstly, the present model does not take into account other levels of speech errors occurring at the syllabic and discourse levels. It is hoped that a segmental level would not only establish the complexity involved in the nature of operations of different modules that are part of the model proposed here but also provide insights that may contribute to accounting for sound organization at the discourse level as well. Secondly, this model has looked only at the word substitutions but not at other frequently occurring phonetic phenomena such as transposition at a segmental level. We have provided a speculative account for the deletion, but we believe a further investigation into the actual mechanisms involved is required. It was felt that the inclusion of other phonological phenomena such as transposition would require further modifications to the present model.

Data availability

The data sets analyzed in the present study are derived from the following sources: Barlow and Gierut's article on "Minimal Pair Approaches to Phonological Remediation." Bernthal, Bankson and Flipson's book titled "Articulation and Phonological Disorders (8th ed)." Sutherland and Gillon's article on "Assessment of phonological representations in children with speech impairment." Bowen's book titled "Children's Speech Sound Disorders."

References

Anthony JL, Aghara RG, Dunkelberger MJ et al. (2011) Approaches to cognition. Harvard University Press, Cambridge

Anthony J, Williams J, Aghara R et al. (2009) Assessment of individual differences in phonological representation. Read Writ 23(8):969–994

Archangeli D, Pulleybank D (1994) Grounded phonology. MIT Press, Cambridge

Backley P (2011) An introduction to element theory. Edinburgh University Press, Edinburgh

Barlow JA, Gierut JA (2002) Minimal pair approaches to phonological remediation. Semin Speech Lang 23(1):57–68

Berent I (2013) The phonological mind. Cambridge University Press, Cambridge

Bernthal J, Bankson N, Flipsen P (2017) Articulation and phonological disorders. Pearson, Boston

Bowen C (2015) Children's speech sound disorders. John Wiley & Sons, West Sussex

Browman CP, Goldstein L (1989) Articulatory gestures as phonological units. Phonology 6(2):201–251

Chomsky N (1975) Reflections on language. Pantheon, New York

Chomsky N, Halle M (1968) The sound pattern of English. Harper & Row, New York

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36(3):181–204

Clark A (2016) Surfing uncertainty: prediction, action, and the embodied mind. Oxford University Press, Oxford

Coleman J (2002) Phonetic representations in the mental lexicon. In: Durand J, Laks B (ed) Phonetics, phonology and cognition. Oxford University Press, Oxford, p. 96

Dell GS (1986) A spreading-activation theory of retrieval in sentence production. Psychol Rev 93:283–321

Dodd B (2014) Differential diagnosis of pediatric speech sound disorder. Curr Dev Disord Rep 1(3):89–196

Dodd B (eds) (2005) Differential diagnosis and treatment of children with speech disorder. Whurr Publications, London

Edwards J, Fourakis M, Beckman ME et al. (1999) Characterizing knowledge deficits in phonological disorders. J Speech Lang Hear Res 42:169–186

Edwards J, Fox RA, Rogers CL (2002) Final consonant discrimination in children: effects of phonological disorder, vocabulary size, and articulatory accuracy. J Speech Lang Hear Res 45:231–242

Ellis AW, Young AW (1996) Human cognitive neuropsychology. Erlbaum, Hove

Flemming E (1995) Perceptual features in phonology. Dissertation, University of California.

Fodor JA (1983) The modularity of mind. MIT Press, Cambridge

Fromkin V (1971) The non-anomalous nature of anomalous utterances. Language 47(1):27–52

Fromkin V (ed) (2012) Speech errors as linguistic evidence. Walter de Gruytar, Berlin

Gafos A (2002) A grammar of gestural coordination. Nat Lang Linguist Theory 20:269–337

Harrikari H (1999) A constraint-based approach to Finnish CV spoonerisms. Nord J Linguist 22(2):111–135

Harris J (1994) English sound structure. Blackwell, Oxford

Hoffman PR, Daniloff RG, Bengoa D et al. (1985) Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. J Speech Lang Hear Res 50:46–53

Humphreys GW, Evett LJ (1985) Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. Behav Brain Sci 8(4):689–740

Idsardi W (2015) Analsis by synthesis techniques. J Acoust Soc Am 137:23–28

Jackendoff R (2002) Foundations of language. Oxford University Press, Oxford

Jun J (1995) Perceptual and articulatory factors in place assimilation: an optimality theoretic approach. Dissertation, University of California.

Kaun A (1995) An optimality-theoretic typology of rounding harmony. Dissertation, University of California.

Kaye J, Lowenstamm J, Vergnaud J (1990) Constituent structure and government in phonology. Phonology 7:193–231

Kaye J, Harris J (1990) Segmental complexity and phonological government. J Linguist 7(2):255–300

Kirchner R (1998) An effort-based approach to consonant lenition. Dissertation, University of California

Leahy J, Dodd B (1987) The development of disordered phonology: a case study. Language Cogn Process 2(2):115–132

Levelt WJ (1989) Speaking: from intention to articulation. MIT Press, Cambridge

Locke J (1985) Phonological acquisition and change. Academic Press, New York

Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Walter Freeman and Co, San Francisco

McNeill BC, Hesketh A (2010) Developmental complexity of the stimuli included in mispronunciation detection tasks. Int J Lang Commun Disord 45(1):72–82

Munson B, Edwards J, Beckman ME (2005) Phonological knowledge in typical and atypical speech-sound development. Top Lang Disord 25:190–206

Namasivayam A, Coleman D, O'Dwyer A et al. (2020) Speech sound disorders in children: an articulatory phonology perspective. Front Psychol. https://doi.org/10.3389/fpsyg.2019.02998

Ohala J (1983) The origin of sound patterns in vocal tract constraints. In: MacNeilage PF (ed) The production of speech. Springer-Verlag, New York, pp. 189–216

Ohala J (1974) Phonetic explanation in phonology. Paper presented at the para-session on natural phonology, Chicago, 18 April.

Ramchand G, Reiss C (2007) The Oxford handbook of linguistic interfaces. Oxford University Press, Oxford

Rvachew S, Ohberg A, Grawburg M et al. (2003) Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. Am J Speech Lang Pathol 12:463–471

Schaefer B, Stackhouse J, Wells B (2016) Phonological awareness development in children with and without spoken language difficulties: a 12-month longitudinal study of German-speaking preschool children. Int J Speech Lang Pathol 19(5):465–475

Shriberg LD, Widder C (1990) Speech and prosody characteristics of adults with mental retardation. J Speech Lang Hear Res 33:627–653

Shriberg LD, Fourakis M, Karlsson HK et al. (2010) Extensions to the speech disorders classification system (SDCS). Clin Linguist Phon 24:795–824

Silverman D (1995) Acoustic transparency and opacity. Dissertation, University of California.

Stackhouse J, Wells B (1997) Children's speech and literacy difficulties I: A psycholinguistic Framework. Whurr, London

Stackhouse J, Pascoe M, Gardner H (2006) Intervention for a child with persisting speech and literacy difficulties: a psycholinguistic approach. Adv Speech Lang Pathol 8(3):231–244

Steriade D (1995) Underspecification and markedness. In: Goldsmith J (ed) The handbook of phonological theory. Blackwell, Oxford, pp. 114–174

Stevens KN, Hanson HM (2010) Articulatory–acoustic relations as the basis of distinctive contrasts. In: Hardcastle WJ, Laver J, Gibbon FE (eds) The handbook of phonetic sciences. Blackwell, Oxford, pp. 424–453

Sutherland D, Gillon GT (2007) Development of phonological representations and phonological awareness in children with speech impairments. Int J Lang Commun Disord 42(2):229–250

Sutherland D (2006) Phonological representations, phonological awareness, and print decoding ability in children with moderate to severe speech impairment. Dissertation, University of Canterbury.

Sutherland D, Gillon GT (2005) Assessment of phonological representations in children with speech impairment. Lang Speech Hear Serv Sch 36:294–307

Sutherland D, Gillon GT (2007) Development of phonological representations and phonological awareness in children with speech impairment. Int J Lang Commun Disord 42:229–250

Terband H, Maassen B, Maas E (2019) A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. Folia Phoniatr Logop 71:216–227

Volenec V, Reiss C (2018) Cognitive phonetics: the transduction of distinctive features at the phonology–phonetics interface. BioL 11:251–294

Walley AC, Metsala JL, Garlock VM (2003) Spoken vocabulary growth: its role in the development of phoneme awareness and early reading ability. Read Writ 16:5–20

Waring R, Knight R (2013) How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. Int J Lang Commun Disord 48(1):25–40

Author contributions

Soujanya Pathi carried out the literature review, collected the data, analyzed it and wrote the major portion of the manuscript. Prakash Mondal conceived the initial idea and

provided significant intellectual inputs to the work described in the paper. Both the authors contributed to the manuscript revisions. Both authors approve of the manuscript and agree to be held accountable for the content presented in the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.P.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.