Journal of Memory and Language 64 (2011) 256–269

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

What models of verbal working memory can learn from phonological theory: Decomposing the phonological similarity effect

Judith Schweppe [a,*], Martine Grice [b], Ralf Rummer [a]

[a] University of Erfurt, Department of Psychology, Germany
[b] University of Cologne, IfL Phonetics, Germany

ARTICLE INFO

Article history:
Received 3 August 2010
revision received 29 November 2010
Available online 7 January 2011

Keywords:
Phonological similarity
Working memory
Phonological features
Auditory perception
Articulation

ABSTRACT

Despite developments in phonology over the last few decades, models of verbal working memory make reference to phoneme-sized phonological units, rather than to the features of which they are composed. This study investigates the influence on short-term retention of such features by comparing the serial recall of lists of syllables with varying types and levels of similarity in their onset consonants. Lists are (a) dissimilar (/fa–na–ga/) (b) acoustically similar (/pa-ta-ka/) or (c) articulatorily similar (/da–la–za/). When no overt articulation is required, we find no decrease in performance for articulatorily similar items as compared to the dissimilar list. However, we are able to show that acoustic similarity clearly impairs recall. It is only when participants recall the lists orally, that performance is impaired for both types of similar lists. These results have implications for accounts of the phonological similarity effect in particular and of verbal working memory in general.

Introduction

Since the 1960s, it has been assumed that short-term storage of verbal information is based on phonological rather than on semantic representations (for an overview, cf. Baddeley, 1986). Moreover, phonemes have been widely considered to be the basic units of verbal short-term memory. Although in the sixties the possibility that verbal short-term memory might be affected by representations below the phoneme level was entertained (Hintzman, 1967; Wickelgren, 1969) and empirical evidence indeed suggested an influence of phonological or phonetic features (Wickelgren, 1965, 1966; see also Brady, Mann, & Schmidt, 1987; Locke & Kutz, 1975; Mueller, Seymour, Kieras, & Meyer, 2003), subphonemic representations in short-term memory have been widely neglected in favour of phonemic representations. Consequently, the nature of subphonemic influences on serial recall tasks has remained an open question: While there seems in principle to be a strong case for the existence of subphonemic influences, it is unclear whether all kinds of features contribute in the same way. For short-term memory models the main question is whether the relevant features are mainly articulatory or mainly acoustic in nature, or both. This is, however, a question that is hard to resolve, since acoustic signals necessarily arise from articulatory movements, and thus neither can be dealt with in isolation. The purpose of the present study is to resurrect this debate by providing deeper insight into the contributions of different types of features, those with a primarily acoustic basis and those with an articulatory basis, drawing upon two major phonological frameworks (*feature geometry*, e.g. Clements, 1992, and *articulatory phonology*, e.g. Browman & Goldstein, 1986).

The phonological similarity effect and how it is explained in working memory theories

The key finding supporting the idea that verbal short-term memory is based on phonological information is the phonological similarity effect (PSE): Lists of items that

* Corresponding author. Address: University of Erfurt, Department of Psychology, P.O. Box 900 221, 99105 Erfurt, Germany.
E-mail address: judith.schweppe@uni-erfurt.de (J. Schweppe).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.11.006

share phonemes in the same positions and syllable structure and that thus sound similar (e.g. the names of letters such as D–C–B–V–P) are recalled more poorly than lists of items that do not (such as the names of the letters R–X–J–T–Q; Conrad & Hull, 1964). This effect is found irrespective of whether items are presented in an auditory or in a visual mode and irrespective of recall modality. The deleterious effect of phonological similarity is also a keystone of the most influential theory of short-term memory, the *working memory model* by Baddeley and Hitch (1974) and its current version (e.g. Baddeley, 2000). Within this model, the effect is traced back to interference based on phoneme overlap in the phonological store, which, in combination with an articulatory rehearsal mechanism, constitutes the phonological loop, a slave system responsible for short-term storage of verbal materials (e.g., Baddeley, Lewis, & Vallar, 1984). The likelihood of confusions is thus greater between items that share one or more phonemes than between dissimilar items. The working memory model is underspecified with respect to the role of subphonemic similarity (e.g., Baddeley & Larsen, 2007). Owing to the influence of the phonological loop model, the phonological store and, consequently, phonemes are widely treated as being the most basic level of verbal short-term memory.

As an alternative to the working memory model, models are currently being promoted that conceive of verbal working memory as "parasitic" (Jones, Macken, & Nicholls, 2004) on (linguistic) processes and representations. Since the PSE is said to be a keystone for models of verbal working memory, these models also have to account for it. In the following, we will introduce two such accounts which differ in their emphasis on peripheral or central processes, focussing on their explanations of the PSE.

Jones et al. (2004); see also Jones, Hughes, & Macken, 2006) have called into question the existence of the entire phonological store concept and have linked the storage of verbal information more closely to the comprehension and production of language in general. Further, they state that they "view short-term retention as parasitic on general perceptual and motor planning processes, particularly (though not solely) as they relate to language" (Jones et al., 2004, p. 670). According to their *perceptual–gestural view of short-term memory*, the PSE can be attributed to two sources, one of them embedded in language *production* and the other embedded in language *perception*. Since in the majority of studies on the PSE items that are treated as phonologically similar are also similar in the way they are articulated, they attribute a substantial part of the PSE to articulatory confusion as it occurs in normal speech production. However, the PSE can also be observed with auditory presentation and written recall and even when participants are engaged in articulatory suppression, an activity which should minimize the influence of articulation-based processes. Jones et al. attribute the persistence of the PSE under these circumstances (i.e. despite disturbance of the articulatory planning process) to *acoustic* similarity and the resulting *perceptual* interference. They argue that under articulatory suppression the PSE is restricted to the initial and final list items. It can thus be reinterpreted in terms of order retention for distinct items benefitting from cues that distinguish the peripheral list positions.

According to Jones et al., an auditory list represents an acoustic object in which the items at boundary positions have a special status. They are either preceded or followed by silence, which makes them particularly distinguishable from the rest of the list. This distinctiveness aids the retention of order for these items. If the items themselves are perceptually similar, the beneficial effect of the additional silence is attenuated. Consequently, the presence of ordering cues (such as silence prior to initial list items and subsequent to final ones) improves recall of perceptually dissimilar but not – or to a lesser degree – recall of perceptually similar items. Thus, the perceptual–gestural view of short-term memory no longer requires a phonological store. Accordingly, Jones et al. (2006, Exp. 2) eliminated the PSE with auditory presentation by engaging participants in articulatory suppression and adding a prefix and a suffix to the list, that is, by hampering articulatory motor planning and acoustic ordering (but see Baddeley & Larsen, 2007, for a critique of these experiments and Jones, Hughes, & Macken, 2007, for a critique of the critique).

In other accounts, sometimes referred to as psycholinguistic short-term memory models, a strong overlap between language and memory models has been promoted as well (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1996; Martin, Lesch, & Bartha, 1999; Martin & Saffran, 1997; Monsell, 1987; Nimmo & Roodenrys, 2004). The common core of these models is that working memory can be conceived of as the activated part of long-term memory (e.g., Cowan, 1999), which for verbal working memory is the language comprehension and production system. These models have been supported by findings that there is more to verbal short-term memory than phoneme-based storage, in particular regarding the influence of lexical and semantic factors on list recall: Recall is better for words than for non-words (Crowder, 1978), for high frequency than for low frequency words (Watkins, 1977), or for concrete than for abstract words (Walker & Hulme, 1999). Beyond the single word level (i.e., in sentence and text recall) even morphosyntatic information influences recall performance (Schweppe & Rummer, 2007; Schweppe, Rummer, & Fürstenberg, 2009). In these accounts, phonological similarity affects serial recall because phonologically similar words (or non-words) are competitors to the items that share more phonemes than phonologically dissimilar competitors. Even though these models have not been developed below the phoneme level, the reference to psycholinguistic and linguistic models suggests a possible extension that also includes an acoustic and articulatory feature level. Then the assumption would be that shared subphonemic input and output representations (i.e., acoustic and articulatory features) cause interference just as shared phonemes do. Specifically, this means that phonemes that share features are more likely to be confused in a short-term memory task. In other words, subphonemic similarity effects should be observed in short-term memory tasks.

Subphonemic similarity effects: first evidence

When looking at phonological similarity as it is generally investigated in short-term memory experiments, the phonological feature level is not addressed (or not addressable). Phonologically similar items in a short-term memory task are usually items with the same syllable structure and the same vowel, varying only the syllable coda (such as /ɛl/–/ɛn/–/ɛs/–/ɛm/–/ɛf/) or onset (/di:/–/si:/–/bi:/–/vi:/–/pi:/), whereas phonologically distinct items involve differences in syllable structure and of one or more sounds at the same time (/ɑ:/–/ɛks/–/ʤeɪ/–/ti:/–/kju:/). Yet, unlike Baddeley's working memory model, the parasitic models described above (Jones' model and psycholinguistic models of short-term memory) address a more basic level. Phonemes themselves can be more or less similar, depending on the number and nature of the features they share. Moreover, certain features tend to be defined in articulatory terms, whereas others tend to have more of an acoustic basis (compare, for instance, the feature inventories of Clements, 1992, and Jakobson, Fant, & Halle, 1952). It should thus be possible to distinguish between similarity effects based solely on shared vowels and syllable structure and those resulting from articulatory or acoustic similarity.

Among the first to demonstrate that the degree of overlap between the features of vowels and of consonants was related to short-term memory confusions was Wickelgren (1965, 1966). He compared different feature systems in terms of their ability to predict the relative frequency of confusions between these sounds in written serial recall. In his study, a combination of the features involving *voicing*, *nasality*, *manner of articulation*, and *place of articulation* made the best predictions for consonant serial recall errors. This rather complex picture might, however, be due to his experimental procedures. For example, in the study reported on in Wickelgren (1966), participants were presented with Consonant–Vowel (CV) syllables but were asked to recall the *consonants*. We cannot thus exclude the possibility of them retaining the items in the form of consonant names, involving a whole range of consonants, vowels and syllable structures.

A similar study was conducted by Ellis (1980). Comparing syllables with the onset consonants /p/, /b/, /m/, /n/, and /s/, he found that those which shared more features tended to be exchanged more frequently in an oral serial recall task than consonants with a lower degree of commonality (this was the case in two thirds of the comparisons).

In a study by Locke and Kutz (1975), associations were observed between children's articulatory impairments (they substituted /w/ for /r/) and their confusion errors in an order reconstruction task when two words were involved that differed only in this respect (ring vs. wing). These errors could not be attributed to misperceptions since the children performed well on perceptual discrimination tasks involving the critical sounds. Such a result provides an indication that articulation may play a role in such a task.

Furthermore, Brady et al. (1987) contrasted serial recall of CV syllables in which consonants shared no feature, one feature, or two features. They also found poorer recall when the consonants shared features than when they did not. Interestingly, however, there was no consistent relationship between the number of shared features and recall performance (i.e., it was not simply the case that the more features are shared, the more exchange errors occur). A closer look at the results reveals that recall was worst when the consonants overlapped in voicing, both when this was the only shared feature and when an additional feature was shared. Even though this aspect of the results was not attended to by the authors themselves, it gives rise to the assumption that similarity is not simply a function of the number of overlapping features but that there may be important qualitative differences across different types of features.

Whenever similarity effects based on feature overlap have been referred to, they have been interpreted as reflecting speech *production* errors, as the errors followed the same principles as slips of the tongue (e.g., Acheson & MacDonald, 2009; Ellis, 1980; Hintzman, 1967; Page, Madge, Cumming, & Norris, 2007). Nonetheless, acoustic and articulatory features have not as yet been varied systematically. In an early paper, Hintzman (1967) did attempt to disentangle auditory from articulatory effects. He demonstrated that confusion errors in short-term serial recall of CVC-syllables preserved place of articulation above chance level, even though in a previous study this dimension had not influenced auditory perception errors in noise (Miller & Nicely, 1955). These findings were, however, controverted soon after. Wickelgren (1969) argued that Hintzman's (1967) results could just as well be explained in terms of auditory or abstract verbal coding, one argument being that the nature of the noise used to obtain errors in an auditory recognition task affects the kind of errors one observes. The question of the relative contribution of acoustic and articulatory coding thus remains unresolved. In the words of Locke and Kutz (1975, p. 187), there "has been substantial controversy as to whether the phonetic effects of speech rehearsal are motor-phonetic or acoustic-phonetic. Specifically, are items confused because they feel alike (Hintzman, 1967; Thomasson, 1970), sound alike (Conrad, 1962), both (Cheng, 1973), or neither (Wickelgren, 1969, posited an "abstract" similarity)?" In order to resurrect this debate, we refer to phonological theories and what they have to say about the role of acoustics and articulation in phoneme similarity.

Acoustic and articulatory similarity: insights from phonological theories

What kinds of features encode articulatory and acoustic similarity? An important articulatory dimension for consonants is the active articulator, that is, that part of the supralaryngeal vocal tract that is actively moved in order to produce the consonant, e.g., the lips, the tip of the tongue, the tongue dorsum and the pharynx (Ladefoged & Maddieson, 1996). In feature geometry terms, this involves the PLACE nodes LABIAL, CORONAL, DORSAL and RADICAL (Clements, 1985; Clements & Hume, 1995; Halle, 1992; Sagey, 1986). Sounds which share an active articulator (and consequently a PLACE node) can thus be considered to be articulatorily similar (see also Hintzman, 1967).

On the other hand, sounds which share the same MANNER of articulation (e.g. whether they are fricatives or plosives, corresponding, all other things being equal, to the positive and negative values of the feature [continuant]), can be regarded as acoustically similar. This is because the effect on the acoustics of manner of articulation is in many cases larger than the effect of place of articulation (see, for example, Fant's, 1973, acoustic feature hierarchy). For instance the acoustic pattern of voiceless plosives involves a silent phase followed by a burst of noise, whereas that of voiceless fricatives involves a continuous random noise pattern, especially in higher frequency regions (Ladefoged, 2001a, 2001b). An example tree in feature geometry is depicted in Fig. 1a.

In phonetic feature systems (i.e. those behind the IPA transcription system), as well as the phonological feature inventories of Chomsky & Halle, 1968, and subsequent feature geometries listed above, it might appear that both PLACE and MANNER are purely articulatorily defined, and thus on an even footing in articulatory terms. However, these features are actually defined in terms which involve "a conflation of articulatory and acoustic properties" (Browman & Goldstein, 1989, p. 222). This observation is not new. In fact there is a long tradition arguing that these features are not solely articulatory (see for example the work of Pike, 1943).

The best established purely articulatory model is articulatory phonology (see e.g. overview papers by Browman & Goldstein, 1986, 1989, 1990, 1992), which incorporates features into a hierarchical structure, referred to as articulatory geometry. In articulatory geometry the traditional PLACE feature is related to different sets of articulators and tract variables. The vocal tract is divided into the oral, glottal and velic systems, and within the oral system the articulators LIPS, TONGUE TIP, TONGUE BODY and TONGUE ROOT are on different tiers. These different place tiers are categorical in nature and are relatively *high up on the feature tree*. By contrast, features which correspond to MANNER are expressed in terms of constriction degree. ("Constriction degree is the analogue within the gestural approach to the manner feature(s)", Browman & Goldstein, 1989, p. 225), the relevant feature for us here being [continuant]. From an articulatory point of view, the constriction is made by a moving set of articulators (related to different articulator tiers) and is not categorical. Constriction degree is a dependent of each articulator node (i.e. from the LIPS node, the TONGUE TIP, BODY or ROOT nodes). This means that constriction degree is dependent on the respective articulator (equivalent to PLACE) node. Differences in articulator (e.g. LIPS vs. TONGUE TIP) are regarded as *major* in articulatory terms, as they involve differences in nodes higher up the tree from which other nodes and features depend. Differences in constriction degree (e.g. full closure vs. critical closure as in stops vs. fricatives) do not involve such a major articulatory distinction (and in fact not even a categorical one, as far as articulatory phonology is concerned). Fig. 1b depicts an example tree in articulatory phonology.

Exactly this issue is taken up by one of the main proponents of feature geometry (Clements, 1992). In feature geometry trees, which Clements acknowledges are mixed ("defined in terms of acoustic, aerodynamic as well as articulatory properties", Clements, 1992, p. 183), manner features (such as continuancy) are represented higher on the tree. He makes it clear that this is not for articulatory reasons but rather due to their phonological behaviour (i.e. their behaviour with regard to phonological statements and processes). Thus, proponents of both articulatory phonology and feature geometry, as well as traditional phoneticians such as Pike, concord with the view that it is not predominantly articulation which contributes to the definition of MANNER features.

Based on the above considerations we take acoustic similarity to correspond to shared MANNER features and articulatory similarity to shared PLACE features. We will now turn to the predictions of the short-term memory theories described above when retention of lists with acoustically similar items, articulatorily similar items, and acoustically and articulatorily distinct items is compared.

Acoustic and articulatory similarity: predictions by theories of working memory

If a subphonemic level is not taken into account, there should be no difference between a list containing maximally distinct consonants and one with consonants which share a number of features and differ in only one feature.

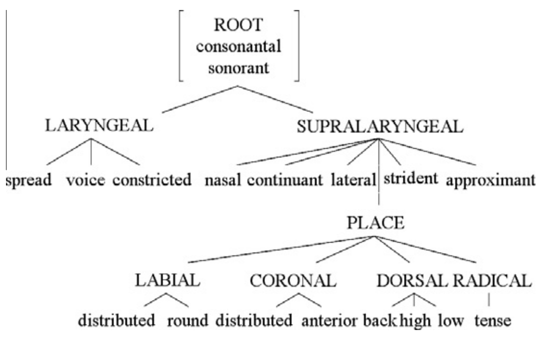

Fig. 1a. *Feature geometry* along the lines of Clements (1985), Sagey (1986) and Halle (1992), adopted from Gussenhoven and Jacobs (2005, p. 163). Note that the MANNER feature [continuant] is higher up the tree than the PLACE nodes.

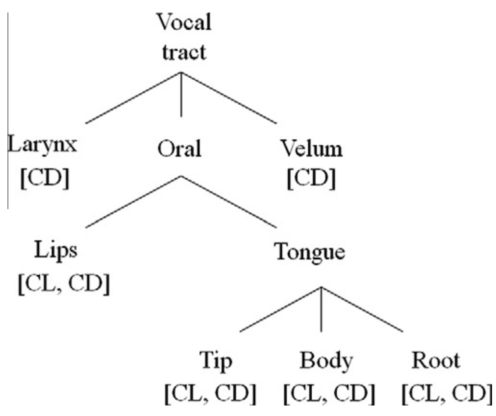

Fig. 1b. Tract variable organization in *Articulatory phonology* according to major articulator, including the parameters constriction location (CL) and constriction degree (CD), adopted from Gafos (1996, p. 12). Note that the MANNER feature CD is below the node for each major articulator (equivalent to PLACE) in the tree.

However, if the acoustic and articulatory feature levels also play a role in verbal short-term memory, as is suggested by both parasitic models introduced above, recall of lists that are characterized by either acoustic or articulatory similarity of the consonants should be worse than recall of lists of consonants that are both articulatorily *and* acoustically distinct. By comparing lists in which either acoustic or articulatory similarity is particularly pronounced and a list in which consonants are both acoustically and articulatorily distinct, the relative contribution of both types of features can be evaluated.

While the two parasitic frameworks predict subphonemic similarity effects, they differ in their predictions regarding the nature of these effects. The perceptual–gestural view (Jones et al., 2004, 2006) explains detrimental effects of acoustic similarity such that acoustic similarity impairs performance because it makes use of distinctive boundary positions less efficiently. The alternative model assigns a potential acoustic similarity effect to interference between acoustic features, irrespective of the use of ordering cues. Consequently, acoustic confusion errors are more likely to occur as more acoustic features are shared. These two views of acoustic influence can be compared by controlling whether or not a prefix and a suffix are added to the list. According to (Jones et al. 2004, 2006), the presentation of a prefix and suffix diminishes the difference between items only when this difference is based on auditory perception. They assume that silence prior to the initial and subsequent to the final list item serves as an ordering cue, the effect of which is reduced if items are perceptually similar.

Consequently, when a prefix and a suffix are presented, neither perceptually dissimilar nor similar lists may benefit from particularly distinctive boundary positions that may serve as ordering cues. Crucially, presentation of a prefix and suffix only eliminates that part of the phonological similarity effect that is due to *acoustic* similarity. With respect to the comparison of distinct, acoustically similar and articulatorily similar lists, this implies that prefix/suffix presentation should reduce the advantages of items that are acoustically distinct as compared to items which are similar in this respect because the former are most affected by this manipulation. If, instead, the acoustic similarity is merely the product of confusions caused by shared phonological features, such an effect should occur irrespectively of prefix/suffix presentation.

In four experiments, we investigate whether, and if so how far, feature overlap can induce interference in a serial recall task. Particularly, we aim to disentangle acoustic and articulatory effects and to test the predictions of the models described above. All experiments vary subphonemic similarity of the list items. In Experiment 1, we use written recall of auditory lists, in the second one written recall of visually presented lists. In the third experiment we investigate oral recall of both auditory and visual lists and in the fourth experiment we compare oral and written recall of auditory lists.

Experiment 1

The experiment was based on a 3 × 2 design with *similarity* (articulatorily similar vs. acoustically similar vs. articulatorily and acoustically distinct) and *prefix/suffix presentation* (without vs. with prefix/suffix) as the two within-subject factors.

Method

Participants

Thirty-six students of Saarland University (all of them native speakers of German) participated in the experiment. They either received course credit or were paid.

Materials

Each trial consisted of five test syllables, which were selected from one of three closed sets of three CV-syllables with articulatorily similar consonants, with acoustically similar consonants, or with distinct consonants. Each set was restricted to three pseudowords, so as to obtain maximal control over featural composition of each list. As a consequence, items were repeated within one trial. An overview of the items and of the consonants' critical acoustic and articulatory features is given in Table 1.

All items shared the same vowel (/a/) and syllable structure (CV) and were in this sense phonologically similar, since we aimed at investigating effects over and above effects merely due to similarities in the syllable structure and vowel. However, the onsets of the syllables were either predominantly articulatorily or acoustically similar, or dissimilar (i.e. neither articulatorily nor acoustically similar). The articulatorily similar lists contained the syllables /da/, /la/, and /za/ where the onsets, /d/, /l/, and /z/, are all produced with the same *active articulator* (tongue tip), in feature geometric terms with a CORONAL place node (Clements & Hume, 1995), at the same *place* of articulation (alveolar). They differ in their manner of articulation (plosive, sonorant, and fricative respectively).

Table 1
The syllables presented in the experiments (acoustically similar lists, articulatorily similar lists, distinct lists, and prefix/suffix) and the critical acoustic and articulatory features for each onset consonant (overlapping manner features indicate acoustic similarity, overlapping place features or place nodes, respectively, indicate articulatory similarity).

	Acoustically similar			Articulatorily similar			Distinct			Prefix/suffix
	pa	ta	ka	da	la	za	fa	na	ga	ja
Manner	Plosive	Plosive	Plosive	Plosive	Sonorant (lateral)	Fricative	Fricative	Sonorant (nasal)	Plosive	Glide
Place node (rel. to active articulator)	LABIAL	CORONAL	DORSAL	CORONAL	CORONAL	CORONAL	LABIAL	CORONAL	DORSAL	CORONAL
Place of articulation	Labial	Alveolar	Velar	Alveolar	Alveolar	Alveolar	Labial	Alveolar	Velar	Palatal

Acoustically similar syllables were /pa/, /ta/, and /ka/. Here, the consonants all had the same *manner* of articulation (plosive), resulting in a similar acoustic signal consisting of closure, release, and aspiration phases. In articulatory terms, /p/, /t/, and /k/ differ in their active articulator and consequently their place of articulation, produced with the lips, tongue tip and tongue body, with labial, alveolar and velar places of articulation respectively. Even though, being plosives, all three consonants involve a closure, the articulatory gestures with which each closure is produced differ considerably. For producing a /p/, the upper and lower lips are pressed together to form a labial closure, for a /t/ the tongue tip is raised to the alveolar region at the front of the palate, whereas for a /k/ the tongue body is raised to the velar region at the back of the palate. In feature geometry the three sounds are represented with different place nodes, LABIAL, CORONAL and DORSAL.

The distinct list items were /fa/, /na/, and /ga/. These three sounds are different in terms of their manner: /f/ is a voiceless fricative, /n/ a voiced nasal, and /g/ a voiced plosive, making them acoustically distinct. In addition to this, they also differ in their active articulator: The articulation of /f/ involves the lower lip, represented with a LABIAL place node, whereas /n/ is articulated with the tip of the tongue (CORONAL), and /g/ with the tongue dorsum (DORSAL).

Stimulus norming. In order to exclude potential differences in the perceived similarity of the acoustically and the articulatorily similar lists, we conducted a rating study. 63 participants who did not take part in any of the experiments were asked to judge how similar they thought each of 30 pairs of syllables to be (on a scale ranging from 1 ("very dissimilar") to 10 ("very similar")). In random order, the three by three pairs included in the experimental lists (/pa–ta/, /pa–ka/, /ta–ka/, /da–la/, /da–za/, /la–za/, /fa–na/, /fa–ga/, and /na–ga/) were embedded in 21 filler items, which also consisted of consonant–/a/ syllables. As expected, the items on the articulatorily similar list were judged as more similar to each other than the items on the distinct list (4.8 vs. 4.3; CI = .17, 1.05), and so were the items on the acoustically similar list (4.9 vs. 4.3; CI = .10, .80). The similarity judgements of the two similar lists, however, did not differ (CI = −.30, .63).

Additionally, the lists did not differ systematically in German biphone and bigram frequencies according to eight different measures along the dimensions of phonological vs. orthographic, lemma based vs. word form based, and type vs. token provided by Hofmann, Stenneken, Conrad, and Jacobs (2007). In two of these comparisons (phonological/lemma based/type and orthographic/lemma based/type), biphone/bigram frequencies were highest for the acoustically similar syllables, in all other cases the highest frequencies were observed for the articulatorily similar syllables.

Procedure

The experiment consisted of 42 trials. One third (i.e. 14) of the five-syllable lists comprised the articulatorily similar items (/da/, /la/, and /za/), one third the acoustically similar items (/pa/, /ta/, /ka/), and one third the dissimilar ones (/fa/, /na/, /ga/). In addition, half of the lists were preceded and followed by the syllable /ja/, which served as a prefix and suffix. This syllable was selected because it shares the same syllable structure and vowel, and at the same time the onset is distinct from the items in all three lists along both the articulatory and acoustic dimensions (the onset of the /ja/ syllable is of a different place of articulation (palatal), and a different manner (glide) from the other list items). Each participant received seven lists per condition. For each list, the items were presented in a pseudorandom order such that each of the three syllables from one set occurred at least once per trial in the respective condition and that a single item was not presented more than twice in succession. To control for an influence of the sequential pattern, the same 14 sequences were used for all three types of lists. The presentation order of the lists was determined at random.

Participants were tested individually. Before the experiment started, they were instructed to listen to the lists carefully and to recall them in the correct order in written form (on an answer sheet) as soon as the final list item had been presented. They were also instructed to ignore the prefix and suffix (/ja/) which would be presented in combination with some of the lists. Prior to the experiment, two practice trials were administered to the participants.

Each trial was preceded by a single tone which lasted 500 ms. 2500 ms later, the first list item was presented via loudspeakers (all syllables were spoken by a male voice and had been recorded in a single session). The duration of all syllables was approximately 500 ms, with an inter-stimulus interval of 500 ms. After the presentation of the final list item, the 15 s recall period started. In this period, participants were required to write down the previously heard list items as they had been presented. The start tone that indicated the beginning of the next list was presented when 12 s of the recall period had passed, that is, 3000 ms before the subsequent trial started. In those trials in which the prefix and suffix were presented, the syllable /ja/ was presented 14 seconds within the recall period (such that the pause between prefix presentation and presentation of the first list item was identical to the inter-stimulus interval, i.e. 500 ms). In these trials, /ja/ was also presented 500 ms after the offset of the final list item (as a suffix). The experiment lasted about 20 min.

Results

We subjected the proportion of correctly recalled items per position to a $3 \times 2 \times 5$ ANOVA with the within-subject factors similarity (articulatorily similar vs. acoustically similar vs. articulatorily and acoustically distinct), prefix/suffix (with vs. without prefix/suffix), and serial position (positions 1–5). Fig. 2 depicts the results.

Most importantly, the analyses revealed a significant main effect for similarity ($F(2, 70) = 14.5$; $MSE = 300.86$; $p < .001$). Recall was poorer for the acoustically similar lists (81.6%) than for both the distinct lists (87.4%; $F(1, 35) = 16.58$; $MSE = 359.47$; $p < .001$; CI = 2.9, 8.7) and the articulatorily similar lists (87.9%; $F(1, 35) = 26.65$; $MSE = 265.47$; $p < .001$; CI = 3.8, 8.8). By contrast, the

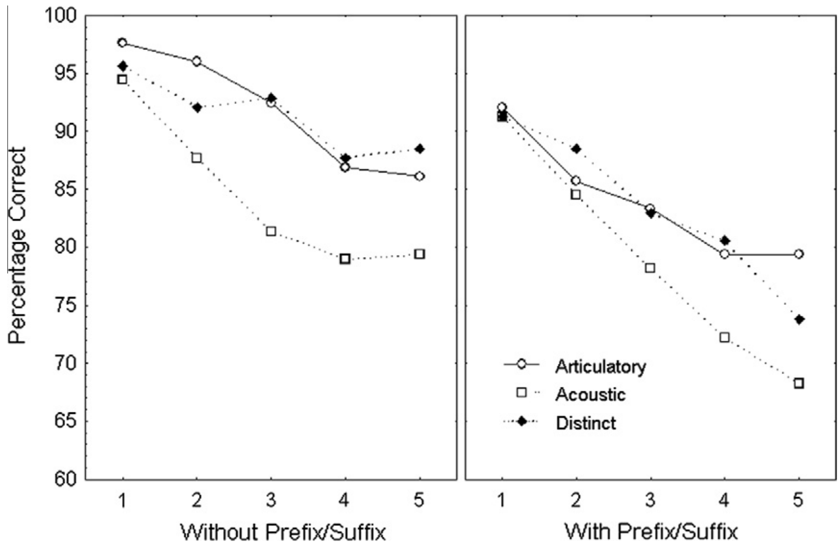

Fig. 2. Percentage of items recalled at their correct serial position as a function of similarity, prefix/suffix presentation, and serial position in Experiment 1. Auditory presentation, written recall.

articulatorily similar (/da-la-za/) and distinct (/fa-na-ga/) lists did not differ from each other ($F < 1$; CI = −2.0, 3.0).

In addition, the main effects for prefix/suffix ($F(1, 35) = 33.33$; $MSE = 407.17$; $p < .001$; CI = 4.6, 9.6) and for serial position ($F(4, 140) = 44.17$; $MSE = 171.60$; $p < .001$) reached significance. Recall performance was worse when a prefix and suffix were presented than when not (82.1% vs. 89.2%) and it declined continuously towards the end of the list (93.7% vs. 89.1% vs. 85.2% vs. 81% vs. 79.2% from position 1 to 5).

The two-way interaction between similarity and prefix/suffix, which would have indicated an influence of the prefix and suffix presentation on the acoustic similarity effect, failed to reach significance. This was the case in both the overall ANOVA ($F < 1$) and when the acoustically similar lists were contrasted with articulatorily and distinct lists combined ($F < 1$). The two-way interactions between prefix/suffix and serial position ($F(4, 140) = 2.89$; $MSE = 110.43$; $p = .025$) and between similarity and serial position reached significance ($F(8, 280) = 2.1$; $MSE = 117.39$; $p = .036$). Both the difference between the conditions with and without prefix/suffix and the difference between the acoustically similar and the distinct list were increased toward the end of the list. As for the acoustic similarity effect, planned comparisons per position revealed significant differences at all positions except the first when collapsing over the two prefix/suffix conditions. The three-way interaction between similarity, prefix/suffix and serial position did not reach significance ($F(8, 280) = 1.47$; $MSE = 114.01$; $p = .167$).

Discussion

Our results provide clear evidence that feature similarity below the phoneme level can have a detrimental effect on serial recall. However, we only found this effect in the case of acoustic similarity (where the MANNER of articulation was shared). Contrary to what was expected on the basis of the perceptual–gestural view, we did not find a substantial reduction of the acoustic similarity effect when a prefix and a suffix were presented.

What is particularly striking is the absence of an articulatory similarity effect. Performance on the distinct list (fa–na–ga) was not better than on the articulatorily similar list (da–la–za). The additional similarity in terms of the active articulator, translated into phonological similarity in terms of the same CORONAL place node (and in fact also the dependent feature [+anterior]), did not impair recall at all. In terms of feature counts, the distinct list was expected to lead to better recall, as the number of feature distinctions was greater.

Before turning to the implications of these findings for the above-described theoretical accounts of the PSE in particular and of verbal working memory in general, a caveat needs to be addressed. One might argue that the higher frequency of acoustic confusion errors is due to errors in perception itself instead of being a memory effect. In other words, the items that sound similar simply might have been misheard. If this were the case, the acoustic similarity effect would be a trivial perceptual phenomenon (perceptual similarity increases perceptual errors) instead of being evidence for the involvement of a more basic representational level in verbal working memory. A straightforward way to avoid the influence of acoustic misperceptions is to apply visual presentation. Moreover, there are further reasons in favour of a visual replication. For one, the first reported observation of "acoustic confusions in immediate memory" by Conrad (1964) was obtained with visually presented items. In addition, one might argue, on the basis of both Baddeley's and Jones et al.'s models, that articulation is more influential with visual than with auditory presentation. According to the phonological loop model, visually presented verbal information enters the phonological store solely via articulatory recoding, while auditory information enters it automatically. Subvocal articulation is thus obligatory with visual verbal items but only optional with auditory ones – at least if the phonological store is to be used for retaining information.

Jones et al. have argued that the main locus of the PSE is articulatory planning. While with auditory presentation there is also an influence of auditory perceptual organization, with visual presentation articulation (or articulatory planning) is the only possible source of a PSE. Thus, both Baddeley's and Jones' model suggest that an influence of articulatory similarity is stronger and therefore more easily detectable with visual than with auditory presentation. With respect to the acoustic similarity effect, the predictions of the perceptual–gestural view and of a framework that assumes all linguistic long-term memory representations to contribute to short-term memory tasks diverge for visual presentation: An acoustic similarity effect which is due to auditory perceptual organization should disappear in the absence of auditory input, whereas an effect that is due to interference between internal representations should persist.

Consequently, we replicated Experiment 1 with visual presentation (dispensing with the prefix/suffix condition), in order to exclude a purely perception-based explanation, to provide more insight into the proclaimed acoustic similarity effect, and to further explore the possibility of an articulatory similarity effect, which had been predicted by both embedded processes models.

Experiment 2

Our second experiment replicates Experiment 1 but with visual presentation instead of auditory representation and without the prefix/suffix factor. It was based on a one-factorial design with *similarity* (articulatorily similar vs. acoustically similar vs. distinct) being varied within subjects.

Method

Participants

Thirty-six students of Saarland University (all of them native speakers of German) participated in the experiment. They either received course credit or were paid. None had taken part in Experiment 1.

Materials

The materials were identical to those in Experiment 1.

Procedure

The procedure matched with that of the first experiment as far as possible. One exception was that all 42 trials were presented without a prefix and a suffix. In addition, all stimuli were presented visually in the centre of a computer screen. Presentation times and inter-stimulus interval were 500 ms, as was the case in Experiment 1.

Results

The proportion of correctly recalled items per position was subjected to a 3 × 5 ANOVA with the within-subject factors similarity (articulatorily similar vs. acoustically similar vs. articulatorily and acoustically distinct) and serial position (positions 1–5). Fig. 3 depicts the results.

As with auditory presentation in Experiment 1, we found a significant main effect for similarity ($F(2, 70) = 11.16$, $MSE = 77.88$; $p < .001$). Recall of the acoustically similar lists (92.6%) was poorer than recall of both the distinct lists (96%; $F(1, 35) = 9.51$; $MSE = 107.63$; $p = .004$; CI = 1.2, 5.6) and the articulatorily similar ones (96.7%; $F(1, 35) = 16.09$; $MSE = 95.27$; $p < .001$; CI = 2.0, 6.2). There was no difference between the articulatorily similar and the distinct lists, however ($F(1, 35) = 1.66$; $MSE = 30.75$; $p = .21$; CI = −.43, 1.94).

In addition, there was a significant main effect for serial position ($F(4, 140) = 26.51$, $MSE = 41.25$; $p < .001$), with recall performance declining from the beginning towards the end of the list (positions 1–5: 98.1% vs. 97.6% vs. 95.8% vs. 94% vs. 90.2%, respectively). The two-way interaction between phonological similarity and serial position did not reach significance ($F(8, 280) = 1.03$, $MSE = 25.68$; $p = .418$).

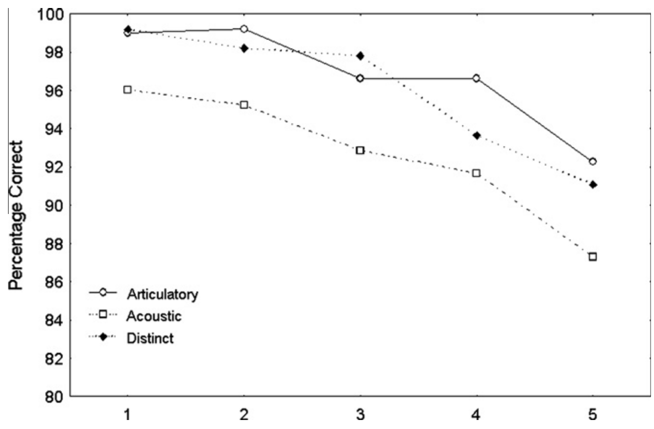

Fig. 3. Percentage of items recalled at their correct serial position as a function of similarity and serial position in Experiment 2. Visual presentation, written recall.

Discussion

As in Experiment 1, distinct lists and articulatorily similar lists involved better recall than acoustically similar lists, while recall performance did not differ between distinct lists and articulatorily similar ones. As pointed out above, the items on the distinct lists differ in PLACE, MANNER and partially in voicing, whereas the items on the articulatory similar lists differ only in MANNER. The manner differences in both of these lists are equivalent in complexity: both lists comprise two obstruents (one fricative and one plosive) and one sonorant. The additional place differences in the distinct lists did not improve recall, as might be expected if a simple count of differences across feature values were involved.

In other words, the data indicate that even in the absence of auditory input consonants which would lead to a similar acoustic signal are more easily confused than consonants that share the same active articulator (and place of articulation) and consonants that are dissimilar along both the acoustic and articulatory dimensions. These data not only replicate the acoustic similarity effect but also give more insight into its origin. Analogously to Conrad's "acoustic confusions in immediate memory", auditory input was not required for acoustic similarity of list items to decrease recall performance. This excludes an alternative explanation in terms of acoustic misperceptions and it contradicts its explanation in terms of auditory perceptual organization.

A puzzling finding is that recall performance was higher in the second experiment (visual presentation) than in the first one (auditory presentation). Usually, auditory lists are recalled better than visual lists. As the comparison is between two experiments that did not only differ in presentation modality (in the auditory experiment half of the lists included a prefix and a suffix), there is, however, good reason to be cautious with an interpretation of the difference in overall recall performance.

In addition to the main purpose of Experiment 2, namely to exclude perceptual errors as the source of the acoustic similarity effect, the use of visual presentation aimed at increasing the role of articulatory processes. A logical next step to further increase the role of articulation is to require participants to overtly articulate the syllables, that is, to use oral instead of written recall. In this case, the role of articulation is not restricted to articulatory planning (in the sense of the perceptual–gestural view) or articulatory rehearsal (in the sense of Baddeley's model) but the recall task cannot be fulfilled without overt articulation. This should increase the influence of articulatory similarity and might, therefore, finally result in an articulatory similarity effect. In order to systematically investigate the unexpected modality differences between Experiments 1 and 2, we have also included presentation modality as an additional within-subject variable.

Experiment 3

This experiment is a modified replication of Experiments 1 and 2 with oral recall. It was based on a two-factorial design with *similarity* (articulatorily similar vs. acoustically similar vs. distinct) and *presentation modality* (auditory vs. visual) being varied within subjects.

Method

Participants

Thirty-six students of the University of Erfurt (all of them native speakers of German) participated in the experiment. They either received course credit or were paid.

Materials

The materials were identical to those in the previous experiments.

Procedure

The procedure matched with that of the first experiments as far as possible. Modality of presentation was blocked and each block consisted of three practice trials and 42 experimental trials. The order of the blocks was counterbalanced across participants: half of the subjects were first presented with the visual lists, the other half started with the auditory lists. Presentation times per item and inter-stimulus interval were 500 ms, as was the case in Experiments 1 and 2. Participants' oral responses were recorded and transcribed.

Results

The proportion of correctly recalled items per position was subjected to a $3 \times 2 \times 5$ ANOVA with the within-subject factors similarity (articulatorily similar vs. acoustically similar vs. distinct), modality (auditory presentation vs. visual presentation), and serial position (positions 1–5). The results are depicted in Fig. 4.

Again, there was a significant main effect for similarity ($F(2, 70) = 44.54$, $MSE = 307.75$; $p < .001$). Recall of the acoustically similar lists (75.91%) was poorer than recall of both the distinct lists (87.96%; $F(1, 35) = 79.18$; $MSE = 270.34$; $p < .001$; CI = 9.2, 14.8) and the articulatorily similar ones (84.27%; $F(1, 35) = 38.87$; $MSE = 323.15$; $p < .001$; CI = 5.7, 11.1). In contrast to Experiments 1 and 2, performance was also better for the distinct lists than for the articulatorily similar ones ($F(1, 35) = 9.07$; $MSE = 270.34$; $p = .005$; CI = 1.2, 6.2).

In addition, there was a significant main effect for serial position ($F(4, 140) = 62.29$, $MSE = 144.89$; $p < .001$), with recall performance declining from the beginning towards the end of the list (positions 1–5: 92.3% vs. 84.8% vs. 82.7% vs. 78.3% vs. 75.5%, respectively). The main effect for modality did not reach significance ($F < 1$). Instead, we observed a significant two-way interaction between similarity and modality ($F(2, 70) = 4.51$; $MSE = 220.24$; $p < .05$), with a marginally significant auditory over visual advantage for the articulatorily similar and the distinct lists (87.4% vs. 84.9%; $F(1, 35) = 3.95$; $MSE = 284.62$; $p = .055$; CI = −.05, 5.05) but a (numerical) visual advantage for the acoustically similar list (74.3% for auditory vs. 77.5% for visual presentation; $F(1, 35) = 2.45$; $MSE = 388.22$; $p = .126$; CI = −.96, 7.47). Also, the difference between the articulatorily similar and the distinct lists was unaffected by presentation modality (3.8% for visual presentation vs. 3.6% for auditory

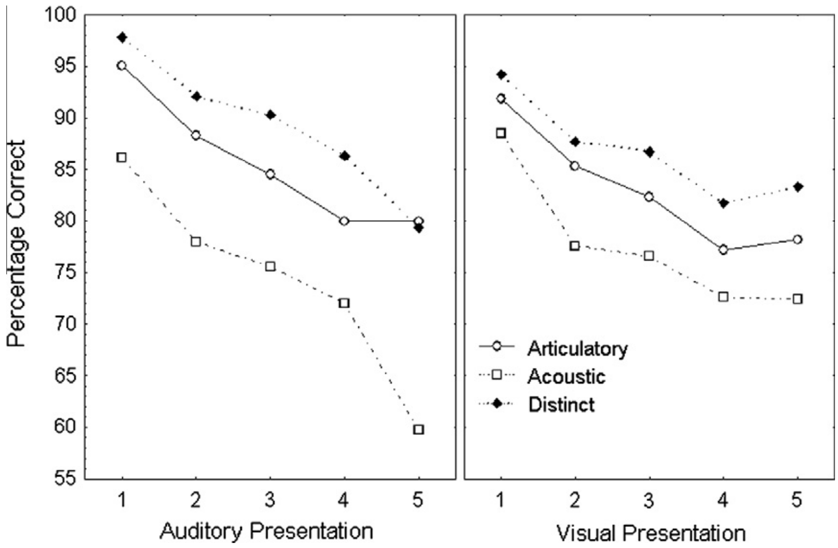

Fig. 4. Percentage of items recalled at their correct serial position as a function of similarity, modality of presentation, and serial position in Experiment 3. Auditory and visual presentation, oral recall.

presentation), while the difference between the acoustically similar and the distinct lists increased from visual to auditory presentation (9.2% vs. 14.9%).

Discussion

The results of Experiment 3 corroborate and extend the findings of the first experiments. As in Experiments 1 and 2, recall of distinct lists and articulatorily similar lists exceeded recall of acoustically similar lists. Unlike the first experiments, however, recall performance for articulatorily similar lists was poorer than for distinct lists. In other words, when recall was oral instead of written, there was an articulatory similarity effect, in addition to an acoustic similarity effect.

Furthermore, the current data shed light on the surprising modality differences observed when comparing Experiments 1 and 2. When presentation modality is systematically manipulated, the picture is clear: As long as the consonants in a list are acoustically (rather) distinct, we find the usual modality effect. In the case of acoustically similar items, however, visually presented lists are recalled better than auditory ones.

The conclusion that articulatory similarity impairs recall only when overt articulation is required has interesting implications beyond our original research question. The comparison between oral and written recall is however a comparison between experiments. As Experiment 3 has shown that the comparison between auditory and visual presentation across Experiments 1 and 2 was misleading, an additional test of the influence of recall modality with auditory presentation within one experiment is thus warranted.

Experiment 4

This experiment provides a within-experiment manipulation of recall modality for auditory lists. It was based on a two-factorial design with *similarity* (articulatorily similar vs. acoustically similar vs. distinct) and *recall modality* (oral vs. written) being varied within subjects.

Method

Participants

Thirty-six students of the University of Erfurt (all of them native speakers of German) participated in the experiment[1]. They either received course credit or were paid.

Materials

The materials were identical to those in the previous experiments.

Procedure

The procedure closely matched with that of the previous experiments. Materials were presented auditorily. Modality of recall was blocked and each block consisted of three practice trials and 42 experimental trials. The order of the blocks was counterbalanced across participants: Half of the subjects had to first recall the lists orally, the other half started with written recall. Presentation times per item and inter-stimulus interval were 500 ms. Participants' oral responses were recorded and transcribed.

Results

The proportion of correctly recalled items per position was subjected to a 3 × 2 × 5 ANOVA with the within-subject factors similarity (articulatorily similar vs. acoustically similar vs. distinct), recall modality (oral vs. written), and serial position (positions 1–5). Fig. 5 depicts the results.

[1] Nine participants had to be excluded since inspection of the audio files revealed that they had not followed the instructions (they either recalled the syllables both orally and in writing in one of the blocks or wrote down the syllables simultaneous to presentation.)

Again, there was a significant main effect for similarity ($F(2, 52) = 41.79$, $MSE = 255.78$; $p < .001$). Recall of the acoustically similar lists (76.22%) was poorer than recall of both the distinct lists (87.67%; $F(1, 26) = 68.11$, $MSE = 260.10$; $p < .001$; CI = 8.60, 14.31) and the articulatorily similar ones (86.46%; $F(1, 26) = 43.23$, $MSE = 327.32$; $p < .001$; CI = 7.04; 13.44). In contrast, performance for the articulatorily similar lists and the distinct lists did not differ ($F(1, 26) = 1.11$, $MSE = 179.91$; $p = .302$; CI = −3.59, 1.16).

In addition, there was a significant main effect for serial position ($F(4, 104) = 59.50$, $MSE = 107.86$; $p < .001$), with recall performance declining from the beginning towards the end of the list (positions 1–5: 91.9% vs. 86.4% vs. 84.0% vs. 79.0% vs. 75.9%, respectively). The main effect for recall modality did not reach significance ($F(1, 26) = 1.30$, $MSE = 495.37$; $p = .265$).

No significant two-way interaction between similarity and recall modality was observed ($F(2, 52) = 1.91$, $MSE = 167.00$; $p = .159$). However, when only the articulatorily similar and the distinct lists were analyzed, the interaction was marginally significant ($F(1, 26) = 3.64$, $MSE = 134.62$; $p = .068$). Planned Comparisons revealed that in written recall there was no difference between the articulatorily similar lists (88.0%) and the distinct lists (87.3%; $F < 1$; CI = −2.84, 4.21). In oral recall, the advantage for distinct (88.0%) over articulatorily similar lists (84.9%) was significant, though ($F(1, 26) = 5.66$, $MSE = 116.18$; $p = .025$; CI = .42, 5.82).

In addition, the two-way interaction between similarity and position reached significance ($F(8, 208) = 2.29$, $MSE = 49.12$; $p = .022$), indicating that the difference between the acoustically similarly lists and the other two list types increased across serial positions (see Fig. 5).

Discussion

The findings of this experiment corroborate the results of the previous ones. The conclusion that articulatory similarity exerted an influence only when overt articulation was required, which was based only on a between experiments comparison, has now been supported by means of a within-subject manipulation. We found an articulatory similarity effect with oral but not with written recall, whereas an acoustic similarity effect was observed in both recall modalities. In line with the first experiments, the acoustic similarity effect was stronger than the articulatory one.

General discussion

By contrasting serial recall of articulatorily similar lists, acoustically similar lists and those which are distinct in both articulatory and acoustic terms, we have demonstrated that acoustic similarity impairs performance no matter if the input is auditory or visual and regardless of whether recall is written or oral. Performance with articulatorily similar items, on the other hand, is equivalent to performance with dissimilar items, whatever the input mode and when recall is written. Only when recall requires overt articulation do we find a disadvantage caused by articulatory similarity. In other words: For articulatory similarity to influence recall performance, articulatory motor movement is essential, while a recall influence of acoustic similarity does not require acoustic-sensory input/auditory perception. Since according to our definitions, acoustic similarity involves the same MANNER of articulation, and articulatory similarity involves the same PLACE of articulation (and the same active articulator), we can conclude that PLACE similarity impairs recall only when recall is overtly articulated.

Interestingly, the dependence of an articulatory similarity effect on overt articulation parallels the findings from two recent speech error studies. Oppenheim and Dell (2008) contrasted errors in overt and inner speech and found that while both overt and inner speech errors exhib-

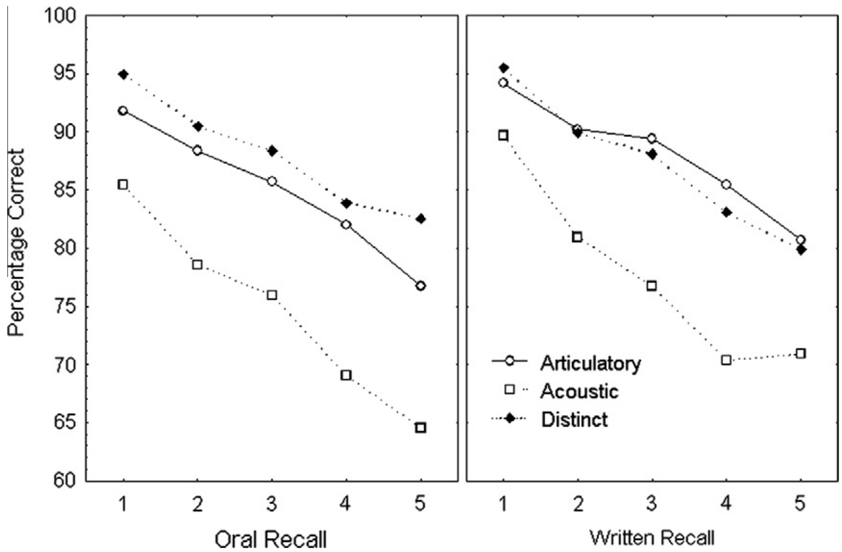

Fig. 5. Percentage of items recalled at their correct serial position as a function of similarity, modality of recall, and serial position in Experiment 4. Auditory presentation, oral and written recall.

ited a lexical bias (i.e., errors that resulted in words were more frequent than errors that resulted in non-words), only in overt slips a phoneme similarity effect (i.e., more errors when the phonemes shared more features) was found. Additionally, when participants had to silently mouth the materials, the error pattern mirrored that of overt speech errors (Oppenheim & Dell, 2010). Oppenheim and Dell (2008, 2010) interpreted their findings such that inner speech is specified at the phoneme level, but not necessarily to the level of articulatory representations and that additional motor execution can create a form of inner speech that incorporates articulatory information. These results resemble our findings that overt oral production is a prerequisite for certain errors to occur. However, there is still a difference between our findings and Oppenheim and Dell's, such that we find a phoneme similarity effect even without motor execution of articulatory gestures – but, crucially, only for shared *acoustic* features. There are two potential reasons for this discrepancy: (1) As the type of feature overlap was not varied in Oppenheim and Dell's (2008, 2010) studies, a similarity effect without motor execution might be found even in their paradigm, when acoustic feature overlap is used and (2) even though the speech error elicitation paradigm used by Oppenheim and Dell (2008, 2010) also included perceptual input and some necessity for retention (as opposed to spontaneous speech errors), our serial recall procedure involved perception and retention to a substantially higher degree. It is thus plausible to assume that the relative importance of certain processes and representations differs between speech errors and short-term memory errors (for a similar argument, see also Schweppe & Rummer, 2007).

A general caveat concerning our findings is the use of lists with repeated items. This was done in order to obtain maximum similarity between the list items. As this is an unusual procedure for serial list recall, it might affect the generalizability of our findings to lists without repetitions. However, as the general effect of subphonemic similarity has been previously observed with more standard lists (e.g., Ellis, 1980; Mueller et al., 2003), it is unlikely that the particular difference between articulatory and acoustic similarity investigated here is restricted to the special case of lists with repeated items. Thus, assuming that our findings provide insight into serial recall beyond lists with repeated items, what are the implications of the present results for the memory models introduced above?

Interpretation in terms of the phonological loop

According to Baddeley, the PSE originates in the phonological store, which is assumed to abstract away from articulatory and acoustic representations. In order to account for the acoustic similarity effect, this model needs to incorporate lower representational levels. Interestingly, following Conrad's (1964) terminology of acoustic confusions in immediate memory, Baddeley (1968) first referred to phonological similarity in terms of acoustic similarity. Later, this was changed to be "relatively neutral with respect to the exact nature of the code" (Baddeley & Larsen, 2007, p. 498). Even though the phonological store concept does not exclude subphonemic similarity effects, they would not have been predicted on the basis of the current version of the model.

Interpretation in terms of the perceptual–gestural view

What about the perceptual–gestural view of short-term memory? Jones et al. (2004, 2006) explicitly attribute the PSE to lower level processes. In terms of the influence of auditory perception, the present data partly favour their view. However, in their model articulatory similarity is supposed to be the main source of the PSE, even with written recall, whereas we only observed an adverse effect of articulatory similarity when overt articulation was required.

A further prediction of Jones' model was that any reduction in performance due to acoustic similarity is attenuated when presentation of a prefix and a suffix makes the boundary positions of a list less distinctive. This was not the case in our first experiment. Suffix presentation impaired performance, particularly at recency, but it did so both for the acoustically distinct lists and for the acoustically similar item lists. Keeping also in mind that the acoustic similarity effect was not restricted to the boundary positions and, what is more, to auditory presentation, the present data call into question its explanation in terms of a boundary effect due to auditory perceptual organization. In particular, the finding of an acoustic similarity effect with visual presentation calls for an alternative account of subphonemic influences on serial recall, since the effect cannot be due to *sensory* acoustic information in this case.

Nonetheless, the finding that the acoustic similarity effect was more pronounced with *acoustic input* in combination with the finding that the articulatory similarity effect was detected only with *articulatory output* clearly indicate a role for sensory-motor codes, as is the core assumption of Jones et al.'s view.

Interpretation in terms of a psycholinguistic view of verbal short-term memory

As an alternative, we have suggested interference at an *acoustic representational* level to be the source of the effect. This is a tacit assumption of psycholinguistic short-term memory models which claim that verbal short-term memory is conditional upon activation of linguistic long-term memory representations (e.g., Martin & Saffran, 1997; Martin et al., 1999; Schweppe & Rummer, 2007). Interestingly, this assumption has not been explicitly claimed for phonological features, probably owing to the focus on higher-level representations. As for the processes operating on these representations, many theories assume interference to be a decisive factor that causes forgetting and have applied this to accommodate the PSE (e.g., Nairne, 1990; Oberauer & Kliegl, 2006). Oberauer and Kliegl (2006, 2010; see also Oberauer & Lange, 2008), for instance, referred to Nairne's (1990) feature model and argued for feature overwriting as the mechanism underlying interference. In their view, items in working memory are represented as vectors with features. For serial recall, each content unit needs to be bound to a context representation, such as list position, by means of synchronous firing. As each feature unit can fire in synchrony with only one context, incoming items can overwrite features that are shared with items already encoded in working memory. Consequently, items in working memory compete for their overlapping feature units and, due to noise in the system, the probability of retrieving a wrong item increases with the number of feature units that have been overwritten. The inclusion of another representational level, and, consequently, of another level at which similarity can cause interference, would allow interference-based theories to account for the acoustic similarity effect as well. So far, this model does not include phonological features but the inclusion of feature representations would be possible, and has already been suggested by Oberauer and Lange (2008).

Among the models of immediate serial recall that have incorporated a model of language *production* is the primacy model (Page & Norris, 1998; see also Page et al., 2007). Here, the PSE is explained in terms of speech errors occurring during a speech output stage which so far includes output representations both at the level of the word as well as at the level of individual speech sounds, or segments. Activation is assumed to flow both top down from word representations to segments and bottom up from segments to words. This bottom-up activation increases the likelihood of errors when the items have segments in common. The model could in principle be extended to include subsegmental featural representations. However, our data have not provided evidence for a pronounced influence of output representations. Nonetheless, the finding of an articulatory effect with oral output is compatible with the primacy model's assumption that speech errors are related to the PSE.

Manner vs. place: why does only one make a difference?

None of the models introduced above would have predicted a feature similarity effect that is stronger for acoustic than for articulatory similarity. Even though both types of parasitic models can account for subphonemic similarity effects, they do not provide an explanation for why acoustic features (or, more precisely, features responsible for the MANNER distinctions) should be subject to overwriting more strongly than articulatory features (or, more precisely, features involving specification of the active articulator and place of articulation, i.e. PLACE).

Interestingly, there is a short-term memory model which emphasizes the importance of similarity at encoding. A key principle of the serial-order-in-a-box (SOB) model (Farrell, 2006; Farrell & Lewandowsky, 2002; Lewandowsky & Farrell, 2008) is *similarity-sensitive encoding*. In SOB, the encoding strength of incoming items is determined by their similarity to the memory content. Items that differ from already encoded information are encoded more strongly than items that are similar to the current contents of memory. This idea corresponds to our finding of a stronger influence of input similarity (related to acoustics) as compared to output similarity (related to articulation).

Additionally, there is an interesting parallel from language processing in the discrepancy between the contribution of input and output features, the *phonological similarity asymmetry* reported by MacKay (1987). According to MacKay (1987, p. 122), "misperceptions involve phonologically similar words more often than misproductions do". He attributes this asymmetry to the greater role of bottom-up priming in language perception than in production, which increases the influence of lower level overlap. Differences in the degree of bottom-up activation spreading to phonemes from input features as opposed to output features might also be a factor contributing to the similarity asymmetry observed here.

In terms of phonological theory, our results concord with the feature geometry laid out in Clements (1992) outlined above. The effect of a MANNER feature, which is attached higher up in a feature tree (e.g. [continuant]), is thus greater than a PLACE node, which is attached lower down.

In sum, we can conclude that we found strong evidence for an influence of (internal) acoustic features on short-term memory, with both auditory and visual presentation and written and oral recall. Furthermore, our data speak for a contribution of sensory-motor codes: Articulatory similarity impairs recall only when overt articulation is required and the impairment due to acoustic similarity is stronger with auditory than with visual lists.

Acknowledgments

This research was supported by the German Research Foundation (DFG) by Grant Ru 891/1-3. Thank you to Philipp von Böselager for producing and processing the test stimuli. We also thank Franca Buge, Kristina Euler, Carla Greving, Carla Gross, Clemens Möckel, Edvina Otto, René Schlegelmilch, Jasmin Schneider, Marie-Kristin Sommer, and Jennifer Spinath for their help in running the experiments and analysing the data. In addition, we are grateful to Robert Hughes, Dylan Jones, Doris Mücke, Dennis Norris, Kathleen Rastle, and three anonymous reviewers for helpful comments.

References

Acheson, D. J., & MacDonald, M. C. (2009). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language, 60*, 329–350.
Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology, 20*, 249–264.
Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.
Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.). *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on evidence for a "perceptual–gestural" alternative. *The Quarterly Journal of Experimental Psychology, 60*, 497–504.
Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology A, 36*, 233–252.
Brady, S., Mann, V., & Schmidt, R. (1987). Errors in short-term memory for good and poor readers. *Memory & Cognition, 15*, 444–453.
Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In: Kingston & Beckman,

(Eds.), Papers in laboratory phonology I. Between the grammar and physics of speech, pp. 341–376.
Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.
Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.
Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.
Cheng, C. (1973). *Acoustic and articulatory coding functions in immediate memory*. Yale: Doctoral dissertation.
Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook, 2*, 225–252.
Clements, G. N. (1992). Phonological primes: Features or gestures? *Phonetica, 49*, 181–193.
Clements, G. N., & Hume, E. (1995). The internal organization of speech sounds. In J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 245–306). Oxford: Basil Blackwell.
Conrad, R. (1962). An association between memory errors and errors due to acoustic masking of speech. *Nature, 193*, 1314–1315.
Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.
Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429–432.
Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.
Crowder, R. G. (1978). Memory for phonologically uniform lists. *Journal of Verbal Learning and Verbal Behavior, 17*, 73–89.
Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1996). A connectionist model of naming errors in aphasia. In J. Reggia, R. Berndt, & E. Ruppin (Eds.), *Neural modeling of cognitive and brain disorders* (pp. 135–156). New York: World Scientific.
Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19*, 624–634.
Fant, G. (1973). *Speech sounds and features*. Cambridge: MIT Press.
Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55*, 587–600.
Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59–79.
Gafos, A. I. (1996). The articulatory basis of locality in phonology. Doctoral dissertation, Johns Hopkins University (published 1999, New York: Garland).
Gussenhoven, C., & Jacobs, H. (2005). *Understanding phonology* (2nd ed.). London: Arnold.
Halle, M. (1992). Phonological features. In W. Bright (Ed.). *International encyclopedia of linguistics* (Vol. 3, pp. 207–212). Oxford: Oxford University Press.
Hintzman, D. L. (1967). Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 6*, 312–316.
Hofmann, M. J., Stenneken, P., Conrad, M., & Jacobs, A. M. (2007). Sublexical frequency measures for orthographic and phonological units in German. *Behavior Research Methods, 39*, 620–629.
Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis. The distinctive features and their correlates*. Cambridge: MIT Press (1969, ninth printing).
Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual–gestural view of short-term memory. *Journal of Memory and Language, 54*, 265–281.
Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *The Quarterly Journal of Experimental Psychology, 60*, 505–511.
Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 656–674.
Ladefoged, P. (2001a). *A course in phonetics* (4th ed.). Boston: Heinle & Heinle.
Ladefoged, P. (2001b). *Vowels and consonants: An introduction to the sounds of languages*. Oxford: Blackwell.
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Wiley-Blackwell.
Lewandowsky, S., & Farrell, S. (2008). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language, 58*, 429–448.
Locke, J. L., & Kutz, K. J. (1975). Memory for speech and speech for memory. *Journal of Speech & Hearing Research, 18*, 176–191.
MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.
Martin, R. C., Lesch, M. F., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 2–39.
Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641–682.
Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338–352.
Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In D. A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading and writing* (pp. 273–311). London: Academic Press.
Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1353–1380.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.
Nimmo, L. M., & Roodenrys, S. (2004). Investigating the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language, 50*, 245–258.
Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*, 601–626.
Oberauer, K., & Kliegl, R. (2010). Interferenz im Arbeitsgedächtnis: Ein formales Modell. [Interference in working memory: A formal model.]. *Psychologische Rundschau, 61*, 33–42.
Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language, 58*, 730–745.
Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528–537.
Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters. *Memory & Cognition, 38*, 1147–1160.
Page, M. P. A., Madge, A., Cumming, N., & Norris, D. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language, 56*, 49–64.
Page, M. P. A., & Norris, D. G. (1998). Modelling immediate serial recall with a localist implementation of the primacy model. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition. Scientific psychology series* (pp. 227–255). Mawah, NJ, US: Lawrence Erlbaum Associates.
Pike, K. L. (1943). *Phonetics*. Ann Arbor: University of Michigan Press.
Sagey, E. (1986). The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT (published by Garland Press, 1990).
Schweppe, J., & Rummer, R. (2007). Shared representations in language processing and verbal short-term memory: The case of grammatical gender. *Journal of Memory and Language, 56*, 336–356.
Schweppe, J., Rummer, R., & Fürstenberg, A. (2009). Beyond sentence boundaries: Grammatical gender information in short-term recall of texts. *Memory & Cognition, 37*, 73–80.
Thomasson, A. (1970). *On the representation of verbal items in short-term memory*. Nijmegen: Drukkerij Schippers.
Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1256–1271.
Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5*, 529–534.
Wickelgren, W. A. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America, 38*, 583–588.
Wickelgren, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America, 39*, 388–398.
Wickelgren, W. A. (1969). Auditory or articulatory coding in verbal short-term memory. *Psychological Review, 76*, 232–235.