Working Memory for Sign Language: A Window Into the Architecture of the Working Memory System

Margaret Wilson
Karen Emmorey
The Salk Institute for Biological Studies

Traditionally, working memory has been divided into two major domains: verbal and visuo-spatial. The verbal domain of working memory can be characterized either by its relationship to language or by its grounding in auditory processing. Because of this ambiguity, languages that are not auditory and vocal (i.e., signed languages) pose a challenge to this conception of working memory. We describe several experiments with deaf users of American Sign Language (ASL) that explore the extent to which the architecture of working memory is determined by the constraints of auditory and visual processing and the extent to which it is determined by the characteristics of language. Various working memory effects were investigated: phonological similarity, word length, and articulatory suppression. The pattern of evidence strongly supports the existence of a sign-based “rehearsal loop” mechanism parallel to the speech-based rehearsal loop. However, we also discuss evidence pointing to differences between the speech loop and the sign loop from forward and backward digit span tasks with deaf and hearing subjects. Despite their similarities based on linguistic properties, the speech loop and the sign loop appear to diverge due to the differing processing demands of audition and vision. Overall, the results suggest that the architecture of working memory is shaped both by the properties of language structure and by the constraints imposed by sensorimotor modality.

Models of working memory typically contain two major components, one used for verbal material, the other used for visuo-spatial material (Baddeley, 1986; Gathercole & Baddeley, 1993; Logie, 1995). The division between these two domains in short-term information processing has been recognized at least since the classic experiments of Brooks (1967, 1968) and Paivio (e.g., Paivio & Csapo, 1969) and appears to reflect a fundamental division in human cognition. It is not clear, though, how best to define this division. On the one hand, the verbal domain can be characterized in terms of language. Thus, printed stimuli (Baddeley, 1986) and lip-read stimuli (Campbell & Dodd, 1980; Gardiner, Gathercole, & Gregg, 1983), although visual, fall primarily within the domain of verbal working memory. On the other hand, the verbal domain can be characterized by its grounding in the auditory and vocal-articulatory properties of speech. Thus, many differences between verbal and visuo-spatial working memory have been attributed to differences between audition and vision.

Because of this ambiguity, languages that are not auditory and vocal, namely signed languages, challenge the traditional model of working memory. Signed languages such as American Sign Language (ASL) are perceived visually and use spatial relationships to convey grammatical information. On these grounds, signed languages seem to fall into the visuo-spatial domain of working memory. On the other hand, signed languages share many characteristics of speech that may be critical for the structure of verbal working memory. On these grounds, signed languages seem to fall into the linguistic domain of working memory. In this article we discuss the theoretical issues about the
nature of working memory that can be addressed by studying sign language and describe a body of research that explores the nature of working memory for sign language.

Why Working Memory for Sign and Speech May Differ

Because sign language and spoken language are grounded in different sensory modalities that have different processing abilities and constraints, there is reason to believe that working memory for sign and for speech may differ systematically. One major difference between auditory and visual perception is their ability to process temporal versus spatial information. For example, iconic memory vanishes within 200–300 milliseconds (Loftus, Duncan, & Gehrig, 1992), while echoic memory persists for two seconds or more (Darwin, Turvey, & Crowder, 1972). Even with nonlinguistic materials, audition is superior in temporal processing: subjects are better at counting rapid sequences of clicks than rapid sequences of flashes (Lechelt, 1975). Numerous findings of this type have led to the conclusion that audition, but not vision, is “intimately tied to time” (Kubovy, 1988, p. 318). Conversely, simultaneous, spatially extended material is coded in vision in ways not possible in audition, given the latter’s poor spatial resolution. Kubovy (1988, p. 318) argues that “space is the province of vision” and that audition is not inherently spatial.

Similar characteristics have been observed in working memory. Memory for the timing of irregular sequences is superior when the stimuli are beeps rather than when they are flashes (Watkins, LeCompte, Elliot, & Fish, 1992). When asked to do free (unordered) recall, subjects tend to preserve the presentation order for speech stimuli but not for print stimuli (e.g., McFarland & Kellas, 1974). Based on such results, Penney (1989) argues that there are “strong unidirectional associations between auditory items” but not between visual items. In contrast, visual working memory seems to benefit when space and simultaneity can be exploited. Several studies have found better recall for visual material with simultaneous presentation than with serial presentation (e.g., Frick, 1985). Penney (1989) concludes that the visual stream of processing within short-term memory “has a stronger spatial component and a relatively weaker temporal one than does the auditory stream.”

Furthermore, speech and sign differ in ways that reflect these differences between the auditory and visual modalities. Signed languages have limited sequential contrasts and permit more simultaneous expression than spoken languages (Emmorey, 1995). ASL can express grammatical information on the face simultaneously with information on the hands, using linguistic facial expressions to mark adverbials, topics, WH questions, conditionals, and relative clauses (Reilly, McIntire, & Bellugi, 1991). At the phonological level, the number of linearly arranged segments appear to be substantially fewer than permitted in spoken languages (Emmorey, 1995; Sandler, 1989). Similarly, at the morphological level, ASL contains relatively few suffixes; instead, ASL expresses morphological distinctions by superimposing movement patterns or spatial distinctions onto the base form of a sign. Signed languages may exhibit fewer linear contrasts because a substantial amount of information can be expressed simultaneously (e.g., handshape, body location, hand/arm orientation, and facial expression can all occur together).

Moreover, when the linear morphology of a spoken language is transferred to the visual modality as Manually Coded English, deaf children exposed to this artificial language spontaneously alter it to create simultaneous (spatial) morphological encoding (Supalla, 1991). This suggests that the visual modality not only easily affords nonlinear structure, but may actually demand it.

Do these differences between sign and speech, and between vision and audition in general, shape linguistic working memory in the two modalities, resulting in differently constructed rehearsal mechanisms? Evidence that this might be so comes from the literature on visuo-spatial working memory (e.g., Logie, 1995) and also from the literature on working memory for printed word stimuli (e.g., Penney, 1989). Working memory in both these domains differs systematically from working memory for speech, in ways that have been attributed to differences between vision and audition. This suggests that working memory for sign language may likewise differ systematically from that for speech.
Why Working Memory for Sign and Speech May Be Similar

However, data from comparisons between speech and meaningless visual stimuli, or even between speech and print, may not be a good basis for making predictions regarding comparisons between speech and sign. Shand and Klima (1981) argue that both speech and sign are “primary language codes,” while print is a “derived code.” We know, for example, that subjects routinely recode print into phonological form in working memory, rather than simply relying on a visual, print-based code. Why should print be such a poor basis for memory as to necessitate this recoding? The answer may lie in the perceptual structure of the stimulus. In terms of perceptual structure, print differs strikingly from dynamically expressed forms of language such as speech and sign, while speech and sign exhibit many similarities.

Signs are not wholistic gestures, but are constructed from a set of meaningless units—handshape, palm orientation, body location, and movement—that are combined in rule-governed ways similar to that at the phonological level in spoken languages (Battison, 1978; Coulter, 1993; Stokoe, 1960). For example, the sublexical structure of ASL exhibits hierarchically organized feature classes, autosegmental representations, deletion and segmentation rules, and a sonority hierarchy (for reviews see Corina & Sandler, 1993; Brentari, 1995). In light of these facts, linguists have broadened the term phonology to refer to the “patterning of the formational units of the expression system of a natural language” (Coulter & Anderson, 1993, p. 5).

Although some aspects of sign phonology are expressed simultaneously, signs also consist of relatively rapid sequences of linearly structured contrasts that unfold in time, just as words do. This linear organization is required in order to contrast minimal pairs and to express certain morphological and phonological rules (e.g., Liddell, 1984). For example, the signs GOOD and ARRIVE are a minimal pair that differs in only one linear segment: the initial location of GOOD is at the chin, whereas the initial location of ARRIVE is near the shoulder. This is comparable to examples like pin, bin in English (see Sandler, 1995). Signs also have syllabic structure (see Wilbur, 1993; Corina & Sandler, 1993; Perlmutter, 1992). Liddell and Johnson (1989) argue that the Hold and Move segments in their model correspond to consonant and vowel segments and that these sequences constitute syllables in both spoken and signed languages. Similarly, Perlmutter (1992, 1993) argues that path movements, by virtue of their perceptual salience, constitute syllable peaks in ASL—a role often played by vowels in English. Further, signs have temporal structuring at the syllabic level, governed by rules similar to those in spoken languages. This temporal structuring involves units of syllabic weight that govern the rhythm of the sign (Perlmutter, 1992; Wilbur, 1993).

This kind of dynamic, rapidly changing perceptual stream, highly structured in its temporal properties, is radically different from most visual stimuli used to test working memory, such as static shapes, pictures, and printed words. Signs are also arguably different in their perceptual properties from other moving stimuli, such as objects undergoing rigid motion, objects undergoing distortion of shape, or even human figures performing nonlinguistic movement.

A further way in which sign and speech are like one another and different from most other stimuli is in the close relationship between perception and production. With sign, as with speech, there is a highly overlearned one-to-one mapping between the perceptual and productive forms. As a consequence, perceived stimuli can be effortlessly translated into their productive form. This permits rehearsal, a relatively automatic quasimotoric process for refreshing information in working memory. Most visual materials used to test working memory do not have this close perception-production link permitting rehearsal. Nonmeaningful stimuli such as shapes and patterns have no canonical motoric representation, while print stimuli, as we have noted, are translated into an entirely different type of code for rehearsal. Rehearsability is critical for the structure of speech-based working memory and may be likewise for sign.

Thus, signed languages offer an opportunity to test the underlying nature of working memory subsystems. To what extent is the architecture of working memory determined by the constraints of sensory processing in particular modalities, and to what extent is it determined by the characteristics of a particular kind of rep-
resentational system such as language? By comparing the organization of working memory in spokenlanguage users and signed-language users, we can explore the relative contributions of language structure and sensorimotor modality to the architecture of working memory.

The Structure of Linguistic Working Memory

In hearing subjects, working memory for language consists in part of a speech-based code that uses the surface form of the language for information maintenance. Evidence for this comes, for instance, from the phonological similarity effect (lists of similar-sounding items are remembered less well than lists of dissimilar items e.g., Conrad & Hull, 1964) and the word length effect (lists of long words are remembered less well than lists of short words, e.g., Baddeley, Thomson, & Buchanan, 1975). These effects indicate that the surface form of the language, with its phonological and articulatory properties, is important for this component of working memory.

The role of surface form is also shown by the fact that memory performance is degraded when subjects perform an irrelevant mouth movement (articulatory suppression, e.g., Murray, 1968) or listen to speech sounds (irrelevant speech, e.g., Salamé & Baddeley, 1982). These effects indicate that memory maintenance uses mechanisms also used in motor planning and perceptual processing and can therefore be disrupted by competition. Thus, the perceptual and productive forms of speech underlie speech maintenance in working memory.

In addition, the interaction of these various effects reveals the structure of this speech-based mechanism, which consists of two components: a phonological store and an articulatory rehearsal process that refreshes information in the store as it fades. Evidence for this two-part structure comes, for instance, from the interaction between phonological similarity and articulatory suppression. When stimuli are auditory (i.e., speech), suppression does not reduce the similarity effect, suggesting that the two effects stem from separate processes or representations. The similarity effect appears to be a product of the phonological store, to which auditory stimuli gain automatic access whether or not the articulatory rehearsal mechanism is available (Baddeley, Lewis, & Vallar, 1984). However, when materials are presented visually, as printed words, suppression eliminates the similarity effect. Apparently, the articulatory mechanism is needed to recode visual materials into phonological form. Further evidence comes from the word length effect. In contrast to the similarity effect, the length effect is eliminated by suppression regardless of presentation mode (Baddeley, Lewis, & Vallar, 1984). Thus, the length effect appears to arise from the articulatory rehearsal process itself and does not occur when that process is unavailable.

Can a similar set of rehearsal mechanisms exist for a visuo-spatial language? To what extent does working memory for sign mirror this “phonological loop” for speech? A high degree of similarity would indicate that the rehearsal loop structure is a function of the type of information being processed, together with a particular kind of expertise on the part of the subject. On the other hand, sign-based working memory may turn out to differ in certain important respects from speech-based working memory. The different processing requirements of the sensori-motor modalities involved may create inherent differences between the two types of language rehearsal systems.

Evidence for a Sign Loop

Growing evidence shows that deaf signers of ASL use an ASL-based memory code for temporary storage, which resembles in some respects the type of speechbased memory code used by hearing subjects (Hanson, 1982; Klima & Bellugi, 1979; Krakow & Hanson, 1985; Poizner, Bellugi, & Tweney, 1981). More recently, we have conducted a systematic inquiry into the structure of this ASL-based memory code, to determine the extent to which it parallels the phonological loop for speech. Our findings indicate a striking degree of resemblance between the two types of memory coding.

The first branch of our investigation concerns the phonological similarity effect and articulatory suppression for sign. For speech, the phonological similarity effect occurs when items to be held in working memory contain many of the same phonemes. Lists with similar items (e.g., mad, man, cad, mat, cap) yield poorer memory performance than lists with dissimilar items
(e.g., pit, day, cow, sup, bar). Articulatory suppression for speech occurs when subjects occupy their vocal apparatuses with an irrelevant activity, such as repeating "ta, ta, ta" during presentation or retention of a list of words. This manipulation reduces memory span because the competing motor activity prevents the use of an articulatory strategy that normally assists memory performance. As noted above, when materials are presented auditorily, suppression does not reduce the phonological similarity effect, indicating that these two effects are independent for speech and may stem from separate processes within working memory.

In our first experiment, reported in full in Wilson and Emmorey (in press-a), we looked for a similar set of effects by testing deaf signers on immediate serial recall of ASL signs. The subjects in this experiment, as well as the other experiments that we review here, were all deaf (hearing loss greater than 80 dB) and were either native ASL signers with deaf families or near-native signers who learned ASL prior to age 6. All subjects considered ASL as their primary language. In this experiment, subjects were shown videotaped sequences of signs that were either similar or dissimilar in terms of handshape (see Figure 1). Subjects were asked to do immediate serial recall under two conditions: in the first condition subjects opened and closed their fists, alternating hands, during the stimulus presentation; in the second condition subjects did not move their hands.

Our results showed a significant phonological similarity effect (worse recall of phonologically similar signs), indicating ASL-based coding in working memory and replicating previous results (e.g., Hanson, 1982; Belleugi, Klima, & Siple, 1975). In addition, the results showed that hand movement also disrupted memory, constituting a manual articulatory suppression effect. Crucially, we found no interaction, indicating that the effects of suppression and similarity are independent of one another. Because video presentation of ASL signs is analogous to auditory presentation of speech (in both cases, no recoding is necessary for storing the material in phonological form), the independence of similarity and suppression parallels the data pattern for the speech-based rehearsal loop. Thus, we can argue that the effects of sign similarity and manual suppression appear to derive from separate components of the memory system.

Our next step was to test similarity and suppression with stimuli that would need to be recoded in order to be stored in phonological form. Recall that, for hearing subjects, when stimuli are not presented as phonological material but rather are presented as print or as nameable pictures, an articulatory process is necessary to recode the stimuli; hence articulatory suppression prevents the similarity effect. In order to test for the parallel of this finding with sign language, subjects were presented with pictures of easily nameable objects. Subjects were asked to produce the ASL sign
for each picture at time of recall. (Pictures were used rather than printed English words, to avoid inducing subjects to use English-based memory.) Pictures were chosen whose corresponding ASL signs were either phonologically similar or dissimilar and were presented either with or without hand motion by the subject.

The results once again paralleled the results with hearing subjects (Wilson & Emmorey, in press-a). The similarity effect occurred when there was no hand motion, indicating that subjects were using an ASL code, but hand motion eliminated the similarity effect. Thus, it appears that an articulatory process is needed to translate materials into an ASL code in working memory. When this articulatory process is unavailable (due to the competing hand motion in the suppression condition), evidence of ASL coding (the phonological similarity effect) disappears.

As in the phonological loop for speech, this pattern of data provides evidence for a buffer that retains information using the phonological structure of signed language and a rehearsal process based on representations for movement of the articulators used in language production, namely the hands. This rehearsal process is used to refresh material in the buffer and to translate material into the phonological code used by the buffer. Recent neuropsychological data suggest that actually two articulatory processes are involved in working memory that can be dissociated from one another: one is used for recoding materials into phonological form, and one used for memory maintenance (e.g., Cubelli & Nichelli, 1992; Fiez, Raife, Balota, Schwartz, Raichle, & Peterson, 1996; Vallar & Cappa, 1987). However, both processes appear to be articulatory in nature, and both are disrupted by articulatory suppression.

To find further support for a sign loop in working memory, we next looked for a sign length effect, analogous to the word length effect for speech in which lists of long words (e.g., hippocampus, tuberculosis, university, Yugoslavia) produce poorer memory performance than lists of short words (e.g., stoat, mumps, school, Greece) (Baddeley, Lewis, & Vallar, 1984). The length effect is presumed to occur because of the involvement of articulatory processes that place a time-limit upon the amount of material that can be retained.

In our study (reported in Wilson & Emmorey, in press-b), we presented "long signs," which cover distance within the signing space or have circular movement, and "short signs," which use short taps or brushes at a single location (see Figure 2).

One concern, however, is that long signs will take longer to report, which may affect performance. That is, information may be disproportionately lost for the long signs before the last item is reported, resulting in reduced performance. Length effects due to report are known to contribute to the size of the word length effect for speech (Avons, Wright, & Pammer, 1994). To avoid this possibility, we used a probe paradigm rather than asking subjects to report the entire list of signs. In the probe paradigm, subjects viewed the list of signs and then saw a single sign. Their task was to report the sign that had occurred immediately after the probe sign in the sequence. This method allows us to test for immediate serial memory while eliminating any effect of report length. Any observed length effect can thus be attributed to the impact of length on retention prior to report.

Our results demonstrated that memory performance was worse for lists with long signs than for the lists with short signs, indicating that sign-based memory is sensitive to the articulation time of signs. Furthermore, subjects performed this task either with or without articulatory suppression. Again, we found
that articulatory suppression reduced performance. However, unlike the case of phonological similarity, suppression eliminated the length effect. That is, memory performance for lists of long and short signs was similar when subjects produced hand motions while viewing the lists to be remembered. When subjects did not move their hands (and thus could rehearse "submanually"), memory performance was worse for lists containing long signs. This pattern of results suggests that the length effect is a direct consequence of articulatory processes that are rendered unavailable under articulatory suppression. Once again, this parallels the findings for speech-based working memory.

Thus, several sources of evidence converge on a model of working memory for sign language. The ASL-based phonological similarity effect, the sign length effect, the manual articulatory suppression effect, and the ways in which interactions between these effects depend upon stimulus presentation mode, all indicate the existence of a phonological loop for sign is highly similar in structure to the phonological loop for speech. This suggests that the phonological loop is a mechanism that develops in response to appropriate linguistic input, regardless of the modality of that input. Our findings suggests that the architecture of working memory is not fixed, but rather responds flexibly to developmental experience.

The model of working memory developed by Baddeley and colleagues (Baddeley, 1986) posits at least two components subordinate to a central executive: the articulatory loop and the visuo-spatial scratch-pad. A study by Reisberg, Rappaport, & O'Shaunessy (1984), in which subjects were taught to use a manual strategy for remembering digits, suggests that the working memory system does not consist of a fixed collection of memory components, but instead can flexibly make use of available coding formats. The research described here demonstrates that not only can the working memory employ such additional formats, but that it can develop a mechanism possessing the same structural details as the articulatory loop. Although the articulatory loop for speech might plausibly be thought to be hardwired, due to the evolutionary history of spoken language, the existence of an articulatory loop for sign cannot be similarly accounted for. Thus, the data discussed here suggest that the components of the working memory system develop in response to appropriate input during development.

Differences Between the Speech Loop and the Sign Loop

Nevertheless, the sign loop and the speech loop may differ in certain important respects, due to the different processing constraints imposed by audition and by vision. As discussed earlier, spatio-temporal coding is a domain in which audition and vision differ, with audition showing relative superiority in temporal coding and vision showing relative superiority in spatial coding (e.g., Kubovy, 1988). If the sign loop and the speech loop are shaped by the perceptual properties of vision and audition respectively, then we might expect to find systematic differences in how serial order coding is accomplished, with speech mechanisms relying more heavily on temporal coding and sign mechanisms relying more heavily on spatial coding.

We know that, for hearing subjects, recalling a list of auditory words in reverse order is more difficult than recalling a list in the order received (Ebbinghaus, 1885/1964), suggesting a unidirectional form of coding, much as time is unidirectional. In contrast to time-based representations, spatial representations do not entail a necessary directionality. A representation that captures serial order information in a spatial form rather than a temporal form might therefore be better suited to reporting a sequence of items backwards. Indeed, it has been shown that when subjects are biased towards supplementing phonological coding with visuo-spatial coding, the difference between forward and backward report diminishes. This has been shown with hearing subjects shown items distributed across space (e.g., Hanson, 1990; Hermelin & O'Connor, 1975); with deaf children who have had inadequate language exposure, and thus might be expected to use a nonlinguistic visuo-spatial strategy (e.g., Blair, 1957; Hermelin & O'Connor, 1975); and even with hearing subjects given ordinary visual presentation of printed words (Powell & Hiatt, 1996). In all these cases, visuo-spatial representation appears to come into play, which in turn appears to facilitate backward report.

Thus, if the visuo-spatial nature of sign language plays a critical role in the representation of ASL in
working memory, then we might expect less difference between forward and backward recall of sign than is found for speech. On the other hand, the critical point may be the properties that sign language has in common with speech, such as its temporal ordering across time or its articulatory properties. The data we have discussed thus far seem to suggest that working memory for sign shares its structure with working memory for speech, and not with working memory for visual materials such as print. If this pattern holds, then we might expect that the reversal of serial order should be difficult for both sign and speech.

To investigate this issue, Wilson, Bettger, Niculae, and Klima (1997) compared deaf children and hearing children on digit span with forward and backward report. The deaf children were native ASL signers and ranged in age from 8 to 10 years; the hearing children were native English speakers within the same age range. Sequences of digits were presented in ASL for the deaf children and in spoken English for the hearing children.

The results indicated that deaf and hearing children showed different patterns of performance on these two tasks (see Figure 3). Deaf children performed equally well on the forward and backward tasks, exhibiting essentially no cost for the requirement to reverse the order of stimulus input. In contrast, hearing children were substantially worse on backward than forward report—the standard finding for spoken-language materials. There was a crossover interaction (hearing were better than deaf on forward report, deaf were better than hearing on backward report), indicating that the different patterns of performance were not due to an overall difference in memory capacity.

These data suggest that sign–based rehearsal mechanisms are not entirely parallel to speech–based mechanisms in working memory. The speech loop appears to be specialized for exact repetition of a sequence of items in the order given, an attribute that may be due to the temporal-processing abilities of audition; whereas the sign loop appears to be less proficient at exact retention but more flexible with respect to ordering, an attribute that may be due to the spatial processing abilities of vision. The form of representation in the sign loop that makes this possible need not be literally visual imagery—it need only be a form of representation that retains at least some of the informational properties of the visual system. The results of this study point to a domain (serial order coding) in which the processing requirements of a particular sensory modality place constraints upon the structure of working memory for a language within that modality.

Taken together, the results discussed in this article highlight both similarities and differences between
working memory for sign and working memory for speech. This pattern of findings indicates that both the universal structural properties of language and the specific processing constraints of sensory modality interact to determine the architecture of working memory. The striking similarities observed between the sign-based and speech-based mechanisms suggest that properties of language (e.g., dynamic, temporally structured sensory input, a close relationship between receptive and productive forms) are sufficient to generate a rehearsal mechanism in working memory. However, this flexibility of the working memory system in developing new components is apparently not limitless. Instead, the exact form of the memory components that develop are constrained by the processing demands of sensory modality.

Notes

1. In fact, a large body of research exists concerning the short-term memory abilities of deaf children and adults. However, the bulk of this work is not aimed at addressing a specifically sign-based component of memory and does not bear directly on the issues raised here.

2. In this study, the hand motion consisted of the 8 handshape for both hands, combined with the movement from the sign WORLD (the two hands circle each other). See Wilson and Emmorey (in press-b) for details regarding this change.

References


Kubovy, M. (1988). Should we resist the seductiveness of the


Working Memory for Sign Language: A Window Into the Architecture of the Working Memory System

Margaret Wilson
Karen Emmorey
The Salk Institute for Biological Studies

Traditionally, working memory has been divided into two major domains: verbal and visuo-spatial. The verbal domain of working memory can be characterized either by its relationship to language or by its grounding in auditory processing. Because of this ambiguity, languages that are not auditory and vocal (i.e., signed languages) pose a challenge to this conception of working memory. We describe several experiments with deaf users of American Sign Language (ASL) that explore the extent to which the architecture of working memory is determined by the constraints of auditory and visual processing and the extent to which it is determined by the characteristics of language. Various working memory effects were investigated: phonological similarity, word length, and articulatory suppression. The pattern of evidence strongly supports the existence of a sign-based “rehearsal loop” mechanism parallel to the speech-based rehearsal loop. However, we also discuss evidence pointing to differences between the speech loop and the sign loop from forward and backward digit span tasks with deaf and hearing subjects. Despite their similarities based on linguistic properties, the speech loop and the sign loop appear to diverge due to the differing processing demands of audition and vision. Overall, the results suggest that the architecture of working memory is shaped both by the properties of language structure and by the constraints imposed by sensorimotor modality.

Models of working memory typically contain two major components, one used for verbal material, the other used for visuo-spatial material (Baddeley, 1986; Gathercole & Baddeley, 1993; Logie, 1995). The division between these two domains in short-term information processing has been recognized at least since the classic experiments of Brooks (1967, 1968) and Paivio (e.g., Paivio & Csapo, 1969) and appears to reflect a fundamental division in human cognition. It is not clear, though, how best to define this division. On the one hand, the verbal domain can be characterized in terms of language. Thus, printed stimuli (Baddeley, 1986) and lip-read stimuli (Campbell & Dodd, 1980; Gardiner, Gathercole, & Gregg, 1983), although visual, fall primarily within the domain of verbal working memory. On the other hand, the verbal domain can be characterized by its grounding in the auditory and vocal-articulatory properties of speech. Thus, many differences between verbal and visuo-spatial working memory have been attributed to differences between audition and vision.

Because of this ambiguity, languages that are not auditory and vocal, namely signed languages, challenge the traditional model of working memory. Signed languages such as American Sign Language (ASL) are perceived visually and use spatial relationships to convey grammatical information. On these grounds, signed languages seem to fall into the visuo-spatial domain of working memory. On the other hand, signed languages share many characteristics of speech that may be critical for the structure of verbal working memory. On these grounds, signed languages seem to fall into the linguistic domain of working memory. In this article we discuss the theoretical issues about the
nature of working memory that can be addressed by studying sign language and describe a body of research that explores the nature of working memory for sign language.

Why Working Memory for Sign and Speech May Differ

Because sign language and spoken language are grounded in different sensory modalities that have different processing abilities and constraints, there is reason to believe that working memory for sign and for speech may differ systematically. One major difference between auditory and visual perception is their ability to process temporal versus spatial information. For example, iconic memory vanishes within 200–300 milliseconds (Loftus, Duncan, & Gehrig, 1992), while echoic memory persists for two seconds or more (Darwin, Turvey, & Crowder, 1972). Even with nonlinguistic materials, audition is superior in temporal processing: subjects are better at counting rapid sequences of clicks than rapid sequences of flashes (Lechelt, 1975). Numerous findings of this type have led to the conclusion that audition, but not vision, is “intimately tied to time” (Kubovy, 1988, p. 318). Conversely, simultaneous, spatially extended material is coded in vision in ways not possible in audition, given the latter’s poor spatial resolution. Kubovy (1988, p. 318) argues that “space is the province of vision” and that audition is not inherently spatial.

Similar characteristics have been observed in working memory. Memory for the timing of irregular sequences is superior when the stimuli are beeps rather than when they are flashes (Watkins, LeCompte, Elliot, & Fish, 1992). When asked to do free (unordered) recall, subjects tend to preserve the presentation order for speech stimuli but not for print stimuli (e.g., McFarland & Kellas, 1974). Based on such results, Penney (1989) argues that there are “strong unidirectional associations between auditory items” but not between visual items. In contrast, visual working memory seems to benefit when space and simultaneity can be exploited. Several studies have found better recall for visual material with simultaneous presentation than with serial presentation (e.g., Frick, 1985). Penney (1989) concludes that the visual stream of processing within short-term memory “has a stronger spatial component and a relatively weaker temporal one than does the auditory stream.”

Furthermore, speech and sign differ in ways that reflect these differences between the auditory and visual modalities. Signed languages have limited sequential contrasts and permit more simultaneous expression than spoken languages (Emmorey, 1995). ASL can express grammatical information on the face simultaneously with information on the hands, using linguistic facial expressions to mark adverbials, topics, WH questions, conditionals, and relative clauses (Reilly, McIntire, & Bellugi, 1991). At the phonological level, the number of linearly arranged segments appear to be substantially fewer than permitted in spoken languages (Emmorey, 1995; Sandler, 1989). Similarly, at the morphological level, ASL contains relatively few suffixes; instead, ASL expresses morphological distinctions by superimposing movement patterns or spatial distinctions onto the base form of a sign. Signed languages may exhibit fewer linear contrasts because a substantial amount of information can be expressed simultaneously (e.g., handshape, body location, hand/arm orientation, and facial expression can all occur together). Moreover, when the linear morphology of a spoken language is transferred to the visual modality as Manually Coded English, deaf children exposed to this artificial language spontaneously alter it to create simultaneous (spatial) morphological encoding (Supalla, 1991). This suggests that the visual modality not only easily affords nonlinear structure, but may actually demand it.

Do these differences between sign and speech, and between vision and audition in general, shape linguistic working memory in the two modalities, resulting in differently constructed rehearsal mechanisms? Evidence that this might be so comes from the literature on visuo-spatial working memory (e.g., Logie, 1995) and also from the literature on working memory for printed word stimuli (e.g., Penney, 1989). Working memory in both these domains differs systematically from working memory for speech, in ways that have been attributed to differences between vision and audition. This suggests that working memory for sign language may likewise differ systematically from that for speech.
Why Working Memory for Sign and Speech May Be Similar

However, data from comparisons between speech and meaningless visual stimuli, or even between speech and print, may not be a good basis for making predictions regarding comparisons between speech and sign. Shand and Klima (1981) argue that both speech and sign are “primary language codes,” while print is a “derived code.” We know, for example, that subjects routinely recode print into phonological form in working memory, rather than simply relying on a visual, print-based code. Why should print be such a poor basis for memory as to necessitate this recoding? The answer may lie in the perceptual structure of the stimulus. In terms of perceptual structure, print differs strikingly from dynamically expressed forms of language such as speech and sign, while speech and sign exhibit many similarities.

Signs are not wholistic gestures, but are constructed from a set of meaningless units—handshape, palm orientation, body location, and movement—that are combined in rule-governed ways similar to that at the phonological level in spoken languages (Battison, 1978; Coulter, 1993; Stokoe, 1960). For example, the sublexical structure of ASL exhibits hierarchically organized feature classes, autosegmental representations, deletion and segmentation rules, and a sonority hierarchy (for reviews see Corina & Sandler, 1993; Brentari, 1995). In light of these facts, linguists have broadened the term phonology to refer to the “patterning of the formational units of the expression system of a natural language” (Coulter & Anderson, 1993, p. 5).

Although some aspects of sign phonology are expressed simultaneously, signs also consist of relatively rapid sequences of linearly structured contrasts that unfold in time, just as words do. This linear organization is required in order to contrast minimal pairs and to express certain morphological and phonological rules (e.g., Liddell, 1984). For example, the signs GOOD and ARRIVE are a minimal pair that differs in only one linear segment: the initial location of GOOD is at the chin, whereas the initial location of ARRIVE is near the shoulder. This is comparable to examples like pin, bin in English (see Sandler, 1995). Signs also have syllabic structure (see Wilbur, 1993; Corina & Sandler, 1993; Perlmuter, 1992). Liddell and Johnson (1989) argue that the Hold and Move segments in their model correspond to consonant and vowel segments and that these sequences constitute syllables in both spoken and signed languages. Similarly, Perlmuter (1992, 1993) argues that path movements, by virtue of their perceptual salience, constitute syllable peaks in ASL—a role often played by vowels in English. Further, signs have temporal structuring at the syllabic level, governed by rules similar to those in spoken languages. This temporal structuring involves units of syllabic weight that govern the rhythm of the sign (Perlmuter, 1992; Wilbur, 1993).

This kind of dynamic, rapidly changing perceptual stream, highly structured in its temporal properties, is radically different from most visual stimuli used to test working memory, such as static shapes, pictures, and printed words. Signs are also arguably different in their perceptual properties from other moving stimuli, such as objects undergoing rigid motion, objects undergoing distortion of shape, or even human figures performing nonlinguistic movement.

A further way in which sign and speech are alike one another and different from most other stimuli is in the close relationship between perception and production. With sign, as with speech, there is a highly overlearned one-to-one mapping between the perceptual and productive forms. As a consequence, perceived stimuli can be effortlessly translated into their productive form. This permits rehearsal, a relatively automatic quasi-motoric process for refreshing information in working memory. Most visual materials used to test working memory do not have this close perception-production link permitting rehearsal. Nonmeaningful stimuli such as shapes and patterns have no canonical motoric representation, while print stimuli, as we have noted, are translated into an entirely different type of code for rehearsal. Rehearsability is critical for the structure of speech-based working memory and may be likewise for sign.

Thus, signed languages offer an opportunity to test the underlying nature of working memory subsystems. To what extent is the architecture of working memory determined by the constraints of sensory processing in particular modalities, and to what extent is it determined by the characteristics of a particular kind of rep-
resentational system such as language? By comparing the organization of working memory in spoken-language users and signed-language users, we can explore the relative contributions of language structure and sensorimotor modality to the architecture of working memory.

The Structure of Linguistic Working Memory

In hearing subjects, working memory for language consists in part of a speech-based code that uses the surface form of the language for information maintenance. Evidence for this comes, for instance, from the phonological similarity effect (lists of similar-sounding items are remembered less well than lists of dissimilar items (e.g., Conrad & Hull, 1964) and the word length effect (lists of long words are remembered less well than lists of short words, e.g., Baddeley, Thomson, & Buchanan, 1975). These effects indicate that the surface form of the language, with its phonological and articulatory properties, is important for this component of working memory.

The role of surface form is also shown by the fact that memory performance is degraded when subjects perform an irrelevant mouth movement (articulatory suppression, e.g., Murray, 1968) or listen to speech sounds (irrelevant speech, e.g., Salamé & Baddeley, 1982). These effects indicate that memory maintenance uses mechanisms also used in motor planning and perceptual processing and can therefore be disrupted by competition. Thus, the perceptual and productive forms of speech underlie speech maintenance in working memory.

In addition, the interaction of these various effects reveals the structure of this speech-based mechanism, which consists of two components: a phonological store and an articulatory rehearsal process that refreshes information in the store as it fades. Evidence for this two-part structure comes, for instance, from the interaction between phonological similarity and articulatory suppression. When stimuli are auditory (i.e., speech), suppression does not reduce the similarity effect, suggesting that the two effects stem from separate processes or representations. The similarity effect appears to be a product of the phonological store, to which auditory stimuli gain automatic access whether or not the articulatory rehearsal mechanism is available (Baddeley, Lewis, & Vallar, 1984). However, when materials are presented visually, as printed words, suppression eliminates the similarity effect. Apparently, the articulatory mechanism is needed to recode visual materials into phonological form. Further evidence comes from the word length effect. In contrast to the similarity effect, the length effect is eliminated by suppression regardless of presentation mode (Baddeley, Lewis, & Vallar, 1984). Thus, the length effect appears to arise from the articulatory rehearsal process itself and does not occur when that process is unavailable.

Can a similar set of rehearsal mechanisms exist for a visuo-spatial language? To what extent does working memory for sign mirror this “phonological loop” for speech? A high degree of similarity would indicate that the rehearsal loop structure is a function of the type of information being processed, together with a particular kind of expertise on the part of the subject. On the other hand, sign-based working memory may turn out to differ in certain important respects from speech-based working memory. The different processing requirements of the sensori-motor modalities involved may create inherent differences between the two types of language rehearsal systems.

Evidence for a Sign Loop

Growing evidence shows that deaf signers of ASL use an ASL-based memory code for temporary storage, which resembles in some respects the type of speech-based memory code used by hearing subjects (Hanson, 1982; Klima & Bellugi, 1979; Krakow & Hanson, 1985; Poizner, Bellugi, & Tweney, 1981). More recently, we have conducted a systematic inquiry into the structure of this ASL-based memory code, to determine the extent to which it parallels the phonological loop for speech. Our findings indicate a striking degree of resemblance between the two types of memory coding.

The first branch of our investigation concerns the phonological similarity effect and articulatory suppression for sign. For speech, the phonological similarity effect occurs when items to be held in working memory contain many of the same phonemes. Lists with similar items (e.g., mad, man, cad, mat, cap) yield poorer memory performance than lists with dissimilar items
(e.g., pit, day, cow, sup, bar). Articulatory suppression for speech occurs when subjects occupy their vocal apparatuses with an irrelevant activity, such as repeating “ta, ta, ta” during presentation or retention of a list of words. This manipulation reduces memory span because the competing motor activity prevents the use of an articulatory strategy that normally assists memory performance. As noted above, when materials are presented auditorily, suppression does not reduce the phonological similarity effect, indicating that these two effects are independent for speech and may stem from separate processes within working memory.

In our first experiment, reported in full in Wilson and Emmorey (in press-a), we looked for a similar set of effects by testing deaf signers on immediate serial recall of ASL signs. The subjects in this experiment, as well as the other experiments that we review here, were all deaf (hearing loss greater than 80 dB) and were either native ASL signers with deaf families or near-native signers who learned ASL prior to age 6. All subjects considered ASL as their primary language. In this experiment, subjects were shown videotaped sequences of signs that were either similar or dissimilar in terms of handshape (see Figure 1). Subjects were asked to do immediate serial recall under two conditions: in the first condition subjects opened and closed their fists, alternating hands, during the stimulus presentation; in the second condition subjects did not move their hands.

Our results showed a significant phonological similarity effect (worse recall of phonologically similar signs), indicating ASL-based coding in working memory and replicating previous results (e.g., Hanson, 1982; Bellugi, Klima, & Siple, 1975). In addition, the results showed that hand movement also disrupted memory, constituting a manual articulatory suppression effect. Crucially, we found no interaction, indicating that the effects of suppression and similarity are independent of one another. Because video presentation of ASL signs is analogous to auditory presentation of speech (in both cases, no recoding is necessary for storing the material in phonological form), the independence of similarity and suppression parallels the data pattern for the speech-based rehearsal loop. Thus, we can argue that the effects of sign similarity and manual suppression appear to derive from separate components of the memory system.

Our next step was to test similarity and suppression with stimuli that would need to be recoded in order to be stored in phonological form. Recall that, for hearing subjects, when stimuli are not presented as phonological material but rather are presented as print or as nameable pictures, an articulatory process is necessary to recode the stimuli; hence articulatory suppression prevents the similarity effect. In order to test for the parallel of this finding with sign language, subjects were presented with pictures of easily nameable objects. Subjects were asked to produce the ASL sign
for each picture at time of recall. (Pictures were used rather than printed English words, to avoid inducing subjects to use English-based memory.) Pictures were chosen whose corresponding ASL signs were either phonologically similar or dissimilar and were presented either with or without hand motion by the subject.

The results once again paralleled the results with hearing subjects (Wilson & Emmorey, in press-a). The similarity effect occurred when there was no hand motion, indicating that subjects were using an ASL code, but hand motion eliminated the similarity effect. Thus, it appears that an articulatory process is needed to translate materials into an ASL code in working memory. When this articulatory process is unavailable (due to the competing hand motion in the suppression condition), evidence of ASL coding (the phonological similarity effect) disappears.

As in the phonological loop for speech, this pattern of data provides evidence for a buffer that retains information using the phonological structure of signed language and a rehearsal process based on representations for movement of the articulators used in language production, namely the hands. This rehearsal process is used to refresh material in the buffer and to translate material into the phonological code used by the buffer. Recent neuropsychological data suggest that actually two articulatory processes are involved in working memory that can be dissociated from one another: one is used for recoding materials into phonological form, and one used for memory maintenance (e.g., Cubelli & Nichelli, 1992; Fiez, Raife, Balota, Schwartz, Raichle, & Peterson, 1996; Vallar & Cappa, 1987). However, both processes appear to be articulatory in nature, and both are disrupted by articulatory suppression.

To find further support for a sign loop in working memory, we next looked for a sign length effect, analogous to the word length effect for speech in which lists of long words (e.g., hippopotamus, tuberculosis, university, Yugoslavia) produce poorer memory performance than lists of short words (e.g., stoat, mumps, school, Greece) (Baddeley, Lewis, & Vallar, 1984). The length effect is presumed to occur because of the involvement of articulatory processes that place a time-limit upon the amount of material that can be retained. In our study (reported in Wilson & Emmorey, in press-b), we presented “long signs,” which cover distance within the signing space or have circular movement, and “short signs,” which use short taps or brushes at a single location (see Figure 2).

One concern, however, is that long signs will take longer to report, which may affect performance. That is, information may be disproportionally lost for the long signs before the last item is reported, resulting in reduced performance. Length effects due to report are known to contribute to the size of the word length effect for speech (Avons, Wright, & Pammer, 1994). To avoid this possibility, we used a probe paradigm rather than asking subjects to report the entire list of signs. In the probe paradigm, subjects viewed the list of signs and then saw a single sign. Their task was to report the sign that had occurred immediately after the probe sign in the sequence. This method allows us to test for immediate serial memory while eliminating any effect of report length. Any observed length effect can thus be attributed to the impact of length on retention prior to report.

Our results demonstrated that memory performance was worse for lists with long signs than for the lists with short signs, indicating that sign-based memory is sensitive to the articulation time of signs. Furthermore, subjects performed this task either with or without articulatory suppression. Again, we found
that articulatory suppression reduced performance. However, unlike the case of phonological similarity, suppression eliminated the length effect. That is, memory performance for lists of long and short signs was similar when subjects produced hand motions while viewing the lists to be remembered. When subjects did not move their hands (and thus could rehearse "submanually"), memory performance was worse for lists containing long signs. This pattern of results suggests that the length effect is a direct consequence of articulatory processes that are rendered unavailable under articulatory suppression. Once again, this parallels the findings for speech-based working memory.

Thus, several sources of evidence converge on a model of working memory for sign language. The ASL-based phonological similarity effect, the sign length effect, the manual articulatory suppression effect, and the ways in which interactions between these effects depend upon stimulus presentation mode, all indicate the existence of a phonological loop for sign is highly similar in structure to the phonological loop for speech. This suggests that the phonological loop is a mechanism that develops in response to appropriate linguistic input, regardless of the modality of that input. Our findings suggest that the architecture of working memory is not fixed, but rather responds flexibly to developmental experience.

The model of working memory developed by Baddeley and colleagues (Baddeley, 1986) posits at least two components subordinate to a central executive: the articulatory loop and the visuo-spatial scratch-pad. A study by Reisberg, Rappaport, & O'Shaunessy (1984), in which subjects were taught to use a manual strategy for remembering digits, suggests that the working memory system does not consist of a fixed collection of memory components, but instead can flexibly make use of available coding formats. The research described here demonstrates that not only can the working memory employ such additional formats, but that it can develop a mechanism possessing the same structural details as the articulatory loop. Although the articulatory loop for speech might plausibly be thought to be hard-wired, due to the evolutionary history of spoken language, the existence of an articulatory loop for sign cannot be similarly accounted for. Thus, the data discussed here suggest that the components of the working memory system develop in response to appropriate input during development.

Differences Between the Speech Loop and the Sign Loop

Nevertheless, the sign loop and the speech loop may differ in certain important respects, due to the different processing constraints imposed by audition and by vision. As discussed earlier, spatio-temporal coding is a domain in which audition and vision differ, with audition showing relative superiority in temporal coding and vision showing relative superiority in spatial coding (e.g., Kubovy, 1988). If the sign loop and the speech loop are shaped by the perceptual properties of vision and audition respectively, then we might expect to find systematic differences in how serial order coding is accomplished, with speech mechanisms relying more heavily on temporal coding and sign mechanisms relying more heavily on spatial coding.

We know that, for hearing subjects, recalling a list of auditory words in reverse order is more difficult than recalling a list in the order received (Ebbinghaus, 1885/1964), suggesting a unidirectional form of coding, much as time is unidirectional. In contrast to time-based representations, spatial representations do not entail a necessary directionality. A representation that captures serial order information in a spatial form rather than a temporal form might therefore be better suited to reporting a sequence of items backwards. Indeed, it has been shown that when subjects are biased towards supplementing phonological coding with visuo-spatial coding, the difference between forward and backward report diminishes. This has been shown with hearing subjects shown items distributed across space (e.g., Hanson, 1990; Hermelin & O'Connor, 1975); with deaf children who have had inadequate language exposure, and thus might be expected to use a nonlinguistic visuo-spatial strategy (e.g., Blair, 1957; Hermelin & O'Connor, 1975); and even with hearing subjects given ordinary visual presentation of printed words (Powell & Hiatt, 1996). In all these cases, visuo-spatial representation appears to come into play, which in turn appears to facilitate backward report.

Thus, if the visuo-spatial nature of sign language plays a critical role in the representation of ASL in
working memory, then we might expect less difference between forward and backward recall of sign than is found for speech. On the other hand, the critical point may be the properties that sign language has in common with speech, such as its temporal ordering across time or its articulatory properties. The data we have discussed thus far seem to suggest that working memory for sign shares its structure with working memory for speech, and not with working memory for visual materials such as print. If this pattern holds, then we might expect that the reversal of serial order should be difficult for both sign and speech.

To investigate this issue, Wilson, Bettger, Niculae, and Klima (1997) compared deaf children and hearing children on digit span with forward and backward report. The deaf children were native ASL signers and ranged in age from 8 to 10 years; the hearing children were native English speakers within the same age range. Sequences of digits were presented in ASL for the deaf children and in spoken English for the hearing children.

The results indicated that deaf and hearing children showed different patterns of performance on these two tasks (see Figure 3). Deaf children performed equally well on the forward and backward tasks, exhibiting essentially no cost for the requirement to reverse the order of stimulus input. In contrast, hearing children were substantially worse on backward than forward report—the standard finding for spoken-language materials. There was a crossover interaction (hearing were better than deaf on forward report, deaf were better than hearing on backward report), indicating that the different patterns of performance were not due to an overall difference in memory capacity.

These data suggest that sign-based rehearsal mechanisms are not entirely parallel to speech-based mechanisms in working memory. The speech loop appears to be specialized for exact repetition of a sequence of items in the order given, an attribute that may be due to the temporal-processing abilities of audition; whereas the sign loop appears to be less proficient at exact retention but more flexible with respect to ordering, an attribute that may be due to the spatial processing abilities of vision. The form of representation in the sign loop that makes this possible need not be literally visual imagery—it need only be a form of representation that retains at least some of the informational properties of the visual system. The results of this study point to a domain (serial order coding) in which the processing requirements of a particular sensory modality place constraints upon the structure of working memory for a language within that modality.

Taken together, the results discussed in this article highlight both similarities and differences between
working memory for sign and working memory for speech. This pattern of findings indicates that both the universal structural properties of language and the specific processing constraints of sensory modality interact to determine the architecture of working memory. The striking similarities observed between the sign-based and speech-based mechanisms suggest that properties of language (e.g., dynamic, temporally structured sensory input, a close relationship between receptive and productive forms) are sufficient to generate a rehearsal mechanism in working memory. However, this flexibility of the working memory system in developing new components is apparently not limitless. Instead, the exact form of the memory components that develop are constrained by the processing demands of sensory modality.

Notes

1. In fact, a large body of research exists concerning the short-term memory abilities of deaf children and adults. However, the bulk of this work is not aimed at addressing a specifically sign-based component of memory and does not bear directly on the issues raised here.

2. In this study, the hand motion consisted of the 8 handshape for both hands, combined with the movement from the sign WORLD (the two hands circle each other). See Wilson and Emmorey (in press-b) for details regarding this change.

References


Kubovy, M. (1988). Should we resist the seductiveness of the


