

# Test of a motor theory of long-term auditory memory

Katrin Schulze<sup>a</sup>, Faraneh Vargha-Khadem<sup>a</sup>, and Mortimer Mishkin<sup>b,1</sup>

<sup>a</sup>Developmental Cognitive Neuroscience Unit, University College London Institute of Child Health, London WC1N 1EH, United Kingdom; and <sup>b</sup>Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, MD 20892

Contributed by Mortimer Mishkin, March 23, 2012 (sent for review July 15, 2011)

Monkeys can easily form lasting central representations of visual and tactile stimuli, yet they seem unable to do the same with sounds. Humans, by contrast, are highly proficient in auditory long-term memory (LTM). These mnemonic differences within and between species raise the question of whether the human ability is supported in some way by speech and language, e.g., through subvocal reproduction of speech sounds and by covert verbal labeling of environmental stimuli. If so, the explanation could be that storing rapidly fluctuating acoustic signals requires assistance from the motor system, which is uniquely organized to chain-link rapid sequences. To test this hypothesis, we compared the ability of normal participants to recognize lists of stimuli that can be easily reproduced, labeled, or both (pseudowords, nonverbal sounds, and words, respectively) versus their ability to recognize a list of stimuli that can be reproduced or labeled only with great difficulty (reversed words, i.e., words played backward). Recognition scores after 5-min delays filled with articulatory-suppression tasks were relatively high (75–80% correct) for all sound types except reversed words; the latter yielded scores that were not far above chance (58% correct), even though these stimuli were discriminated nearly perfectly when presented as reversed-word pairs at short intrapair intervals. The combined results provide preliminary support for the hypothesis that participation of the oromotor system may be essential for laying down the memory of speech sounds and, indeed, that speech and auditory memory may be so critically dependent on each other that they had to coevolve.

evolution | mimic | arcuate fasciculus

The proficiency with which monkeys perform tests of both visual and tactile recognition does not extend to auditory recognition. In vision and touch, monkeys master the rule for one-trial recognition memory extremely rapidly, within several daily sessions (1, 2); and once they have learned the rule, it can be shown that they have stimulus–retention thresholds (performance at 75% accuracy) of 10–20 min after viewing or palpating a novel stimulus for only 1–2 s (3, 4). In audition, by contrast, monkeys acquire the rule for one-trial memory exceedingly slowly, requiring a full year or two of training before they can master it, if they succeed at all; and if they do succeed, their stimulus–retention thresholds are found to extend no longer than 30–40 s after stimulus presentation (5). This marked disparity in mnemonic ability across sensory modalities suggests that, in audition alone, monkeys seem unable to store stimulus representations in long-term memory (LTM) and, consequently, appear to be limited mnemonically to the time period covered by short-term memory. Humans, on the other hand, are highly proficient at storing lasting representations of auditory stimuli, such as words and tunes, thereby enabling their later recognition. What accounts for these striking mnemonic differences between audition and other sensory modalities in the monkey, and between audition in the monkey and audition in humans?

In the present study, we tested a hypothesis derived from the following considerations. Whereas trial-unique visual and tactile stimuli used in recognition memory tasks are commonly presented as stationary items, most trial-unique sounds used in such tasks fluctuate at high, millisecond speeds. Perhaps unlike stimuli that are stationary, stimuli that fluctuate rapidly cannot be

packaged for storage in the relevant sensory processing system alone. Such packaging or integration of a fluctuating stimulus, even within a very short temporal window, may require the assistance of the motor system, which seems uniquely organized to chain link rapid sequences. Unlike monkeys, humans have a dense and complex pathway connecting the auditory system in the posterior temporal region with the oromotor system in the ventrolateral frontal region; this pathway, the arcuate fasciculus (AF) (e.g., refs. 6–8), can often transform an acoustic sequence into a subvocal oromotor sequence, as evidenced by the listener's ability to vocally reproduce or mimic the sound (9).<sup>\*</sup> This integrated acoustic/oromotor sequence might constitute the stored central representation of that sound, enabling its later recognition. Alternatively, if the acoustic stimulus cannot be easily mimicked—which can occur, for example, with certain environmental sounds—it can often be tagged with a label, either a name or an already stored representation of a visual or other nonacoustic sensory stimulus; this newly associated label could then serve as a mnemonic surrogate and obviate the need for storing the representation of the sound, per se.

A corollary hypothesis derived from the foregoing considerations is that an acoustic stimulus that can be neither mimicked nor labeled cannot be stored for subsequent recognition. We tested this hypothesis by comparing the ability of participants to recognize sounds of different types that varied widely in the degree to which they could be reproduced or labeled. Each stimulus type (Words, Pseudowords, Nonverbal Sounds, and Reversed Words) was presented as a study list, and this list was followed after a 5-min delay by a test list requiring old/new judgments. Pairs of reversed words were also presented later in a separate, same/different discrimination test.

## Results

As illustrated in Fig. 1, recognition scores were highest for Words (80.8 ± 2.0% correct responses), somewhat lower for both Pseudowords (76.4 ± 1.95) and Nonverbal Sounds (74.7 ± 1.9), and lowest for Reversed Words (57.5 ± 2.5), but significantly above chance (one-sample *t* test ( $t_{[31]} = 3.02, P < 0.005$ ). An ANOVA with stimulus type as the within-subject factor showed a highly significant main effect ( $F_{[3,29]} = 28.99, P < 0.0001$ ). Although post hoc paired comparisons (Bonferroni corrected  $\alpha$ -level:  $P = 0.008$ ) between Words and Pseudowords ( $t_{[31]} = 1.90, P = 0.07$ ), between Words and Nonverbal Sounds ( $t_{[31]} = 2.49, P = 0.02$ ), and between Pseudowords and Nonverbal Sounds ( $t_{[31]} = 0.78, P = 0.44$ ) failed to reach significance, comparisons between Reversed Words and each of the three other stimulus types were highly significant (Words:  $t_{[31]} = 7.74, P < 0.0001$ ; Pseudowords:  $t_{[31]} = 6.60, P < 0.0001$ ; and Nonverbal Sounds:  $t_{[31]} = 5.48, P < 0.0001$ ).

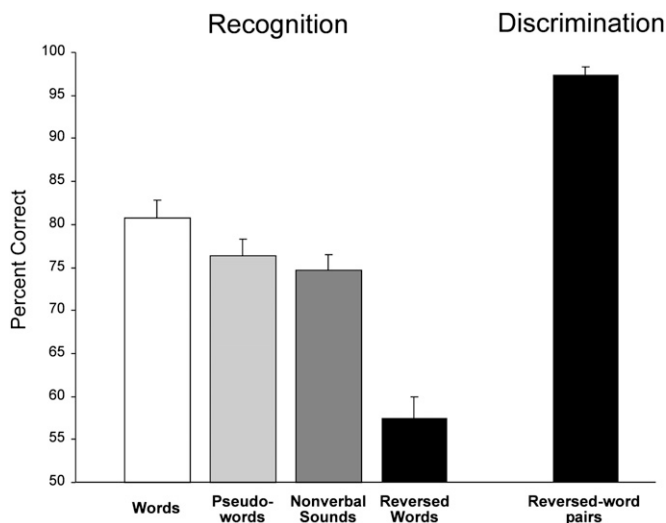
Author contributions: K.S., F.V.-K., and M.M. designed research; K.S. performed research; K.S. analyzed data; and K.S., F.V.-K., and M.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: mishkinm@mail.nih.gov.

<sup>\*</sup>Unlike the arcuate fasciculus, a different temporofrontal pathway, this one coursing through the extreme capsule to link the middle section of temporal cortex with the ventrolateral and dorsolateral frontal regions, is highly developed not only in humans but in nonhuman primates as well (8, 10, 11).



**Fig. 1.** Percentage of correct responses on the auditory recognition and auditory discrimination tasks, considered to be measures of memory and perception, respectively. Recognition of each of the four stimulus types was tested separately after 5-min study-test delays filled with articulatory-suppression tasks; same/different discrimination of the reversed-word pairs was tested with intrapair delays of 0.5 s (error bars indicate SEM). Mean recognition score on reversed words fell significantly below the score on each of the other stimulus types (all  $P$  values  $<0.0001$ ).

Because a Shapiro–Wilk test indicated that the distribution of recognition scores for Pseudowords ( $P = 0.014$ ) was significantly abnormal, we performed additional, nonparametric analyses, and these analyses confirmed the results of the ANOVA and of each post hoc comparison. Thus, a Friedman test with stimulus type as the within-subject factor was highly significant [ $\chi^2(3, n = 32) = 40.3, P < 0.0001$ ] and the significant post hoc comparisons (Wilcoxon signed rank tests, Bonferroni corrected  $\alpha$ -level:  $P = 0.008$ ) were those between reversed words and each of the three other stimulus types (Words, Pseudowords, and Nonverbal Sounds,  $Z$ s of 4.59, 4.45, and 4.15, respectively, all  $P$  values  $<0.0001$ ).

Recognition performance was also analyzed using the bias-free measure of sensitivity,  $d'$ , of signal detection theory by comparing hits (i.e., number of correct responses for new items/number of new items) and false alarms (i.e., number of incorrect responses for old items/number of old items). This index was highest for Words ( $2.18 \pm 0.2$ ), somewhat lower for both Pseudowords ( $1.72 \pm 0.17$ ) and Nonverbal Sounds ( $1.54 \pm 0.15$ ), and lowest for Reversed Words ( $0.39 \pm 0.14$ ). Because the distribution of the performance data was not completely normal (for Pseudowords:  $P = 0.017$  and for Nonverbal Sounds:  $P = 0.021$ ; Shapiro–Wilk test), the nonparametric Friedman test was carried out. This test with stimulus type as the within-subject factor showed a highly significant main effect [ $\chi^2(3, n = 32) = 44.1, P < 0.0001$ ]. Post hoc paired comparisons (Wilcoxon signed rank test, Bonferroni corrected  $\alpha$ -level:  $P = 0.008$ ) for reversed words and each of the three other stimulus types were highly significant (words:  $Z = 4.86, P < 0.0001$ ; Pseudowords:  $Z = 4.34, P < 0.0001$ ; and Nonverbal Sounds:  $Z = 4.23, P < 0.0001$ ). In addition, Words and Nonverbal Sounds also differed significantly ( $Z = 2.77, P = 0.006$ ).

The difficulty in recognizing Reversed Words does not appear to be attributable to a failure of sensory processing, inasmuch as the participants scored  $97.4 \pm 0.93\%$  correct in distinguishing between the members of reversed-word pairs, when the paired stimuli were separated by 0.5-s intervals in the auditory discrimination task (Fig. 1).

To determine whether there was any relationship between scores on the different stimulus types, or between scores on

a particular stimulus type and on discrimination of reversed-word pairs, we calculated bivariate correlation coefficients. No significant correlations were observed (all  $P$  values  $>0.065$  before correction for multiple comparisons).

A mixed-type ANOVA performed with stimulus type as the within-subject factor and language (native vs. nonnative English speakers) as a between-subject factor confirmed the main effect of stimulus type ( $F_{[3,28]} = 29.22, P < 0.0001$ ), but yielded neither a main effect of language ( $F_{[3,28]} = 0.175, P = 0.68$ ) nor an interaction between language and stimulus type ( $F_{[3,28]} = 1.25, P = 0.30$ ).

The other variables we investigated likewise failed to affect auditory recognition memory. Specifically, with regard to the different articulatory-suppression filler tasks, subvocal counting of tones had no greater effect than subvocal counting of visual symbols on the later recognition of any type of auditory stimulus (all  $P$  values  $>0.26$ ). Also, there were no differences between males and females or between participants with and without experience playing a musical instrument, in the recognition of any stimulus type (all  $P$  values  $>0.13$ ), or on the discrimination of reversed-word pairs (both  $P$  values  $>0.65$ ).

## Discussion

We investigated the potential contribution to human auditory recognition memory of the degree to which acoustic signals can be mimicked or labeled, by using sounds that differed widely along these two dimensions. Recognition scores after 5-min delays filled with articulatory-suppression tasks were highest for Words (81%), somewhat lower for both Pseudowords and Nonverbal Sounds (76 and 75%, respectively), and lowest by far for Reversed Words (i.e., the Words played backward, 58%). We also found that the difficulty the participants had in recognizing the Reversed Words was not due to a failure in stimulus perception, because they performed nearly perfectly in discriminating reversed-word pairs when the within-pair delay was limited to 0.5 s. Our results thus provide preliminary support for the hypothesis that humans cannot perform one-trial recognition of novel auditory stimuli that they can neither reproduce nor label.

Many more tests of this proposal need to be conducted. For example, just as intensive exposure of an adult learner to the atypical word forms of an unfamiliar language can lead to success in recognizing the words of that language, perhaps familiarizing a study participant with an extensive list of reversed words might in time improve the memorability of a reversed-word study list. However, if it does, and if, despite the familiarization, those stimuli should remain unpronounceable, then according to our hypothesis the improvement could come about through the association of a label with each reversed word, just as with any other unmimickable nonverbal sound (see below).

However, what is the neural circuitry that enables one-trial recognition of auditory stimuli when fluctuating acoustic signals *can* be either mimicked or labeled? Before elaborating on the proposals outlined earlier, we must note first that, if our hypothesis is correct, each of the three stimulus lists on which the participants performed relatively well—Words, Pseudowords, and Nonverbal Sounds—would have activated partly different circuits. Although none of these proposed circuits has yet been worked out in any detail, some of the major structures and interconnections that would be needed to recognize Words and Nonverbal Sounds can be surmised from the available evidence. For example, recognition of the familiar Words would require only that their previously stored representations be reactivated at test and identified as having recently been activated during presentation of the study list; presumably, this form of episodic memory would depend on the interconnections of the superior temporal auditory processing stream (12–14), the lateral temporal semantic system (14, 15), and the medial temporal lobe, including the hippocampus (16–20).

By contrast, successful “recognition” of the Nonverbal Sounds would presumably require reactivation at test of previously associated verbal and/or visual tags for some of the recently presented study-list sounds and newly associated tags for others of those sounds, thereby using much the same temporal lobe circuitry as that used for words (21) except for the substitution of cross-modal labeling—e.g., sound/name or sound/visual image—for the sound itself. This type of memory would therefore not be acoustic recognition in the strict sense, in that the centrally stored representation that carries the memory would not be the sound, per se, but its retrieved associate, which could often be the stored representation of a stimulus in another sensory modality.

Still another mechanism would be needed, however, to account for recognition of the pronounceable but unfamiliar Pseudowords. As indicated at the outset, we hypothesized that one way in which one-trial auditory recognition of novel speech sounds might occur is through the automatic transformation of the acoustic sequence into a subvocal oromotor sequence, and that this integrated acoustic/oromotor signal could then be stored as a lasting central representation. Such automatic transformation and integration would of course require a strong bidirectional link between the temporoparietal auditory system and the ventrolateral frontal oromotor system, and, indeed, just such a link is provided by the arcuate fasciculus (e.g., refs. 6, 22). Once encoded and stored in the auditory processing stream during study, the central representation of the Pseudoword would enable auditory recognition at test following the same sequence of events and using the same circuits as those used for Words, i.e., short-circuiting the arcuate fasciculus, or at least not requiring its reactivation. Of course, unlike a Word, the recognized Pseudoword would have no meaning until it was associated with the central representation of another stimulus; at that point it would be recognized in exactly the same way as a Word, through direct activation of its stored representation in the auditory system, independent of the auditory–oromotor connection.

**Motor Theory of Speech Perception.** Invoking the arcuate fasciculus in support of our proposal raises the issue of what role this tract plays in speech perception. The notion that speech production and processing speech sounds are intimately linked was proposed long ago (23, 24). However, the idea was introduced at that time in an attempt to explain, not long-term auditory memory, but the perception of speech. In its most recent version (24), this theory proposed that speech is perceived as “specific patterns of intended (oromotor) gestures” and that speech perception is possible because listeners do not hear speech as ordinary sounds; rather, they use the relation between acoustic signal and oromotor gesture to perceive speech.

Although the motor theory of speech perception lay dormant for several years, it recently received renewed attention with the discovery in monkeys of “mirror neurons,” namely, neurons in the ventral premotor cortex (area F5) that discharged not only when the monkey made a visually guided movement, such as reaching for an object and grasping it, but also when the monkey observed the experimenter perform the same movement (25, 26). *Audiovisual* mirror neurons were soon observed as well (27), this type of cell having been uncovered in monkeys in a more ventral part of area F5 (28); these neurons were activated not only when the subject’s own action (e.g., dropping a stick) resulted in a sound but also when the subject heard the same sound produced by the same, but unseen, action of another. By providing a neural correlate for a link between perception and action, discovery of the mirror-neuron system initiated a new wave of research on the motor theory of speech perception (e.g., ref. 29), uncovering evidence of activation in oromotor-related areas while listening to speech (30–32) and of interference with the

perception of speech by transcranial magnetic stimulation (TMS) of premotor and motor cortex (33–35).

Although these new findings indicate that the oromotor system plays a role in speech processing, the motor theory of speech perception in its strong form—which should predict perception failure in the absence of oromotor activation—remains widely challenged (for an overview see 36, 37–39), and particularly so by evidence that speech perception is preserved in “Broca’s aphasics” (40). However, whether or not oromotor activation contributes to the perception of speech, it appears to play little or no role in the perception of reversed words, inasmuch as these were perceived well enough to be easily discriminated from each other when they were presented in pairs separated by 0.5-s intrapair intervals.

**Motor Theory of Speech Recognition.** At the same time, the participants’ extremely poor recognition of the nonmimickable Reversed Words supports our hypothesis that a novel speech sound’s pronounceability—i.e., its potential to activate the speech production system automatically and subvocally, perhaps via an auditory mirror-neuron system—may well be essential for laying down a lasting representation of that sound. The neural mechanism is unknown, but, to elaborate on our speculative proposal, this mechanism could involve a multistep process mediated in large part by the AF, in which: (i) signals in temporoparietal cortex produced by processing a novel, fluctuating speech sound would be transmitted via the AF to the ventrolateral frontal cortex and, from there, via corticostriatal pathways (41), signals can be transmitted to the striatum for mapping onto representations of a sequence of articulation and coarticulation plans for reproducing that sound (39, 42); (ii) signals representing this covert, oromotor sequence would then be fed back to the temporoparietal cortex via the AF’s reciprocal projections (43, 44); and (iii) these repackaged signals representing the newly integrated auditory–oromotor sequence would be processed by the lateral temporal auditory stream and transmitted to the medial temporal lobe, where they would be initially encoded and stored. This proposal of a motor theory for speech recognition is similar in some respects to one proposed for vocal learning by songbirds (45, 46), including the critical step (ii) above) of an oromotor feedback or corollary discharge to auditory cortex, thereby establishing a precise sensorimotor correspondence between audition and vocalization.

A potential objection to a motor theory of recognition memory for new words is that some patients with severe problems in speech production are nonetheless capable not only of speech comprehension but also of learning language through reception alone (47, 48). Although, to our knowledge, such patients have not been tested on verbal LTM tasks of the type used here, studies of verbal working memory (WM) indicate that the outcome in verbal LTM may depend on the neural basis of the speech impairment. Thus, whereas either congenital or acquired anarthria, a neuromotor disorder that disrupts the control of muscles required for speech (49), can leave verbal WM largely intact (50, 51), *apraxia* of speech, a higher-level disorder in which speech production impairment results from a disorder of speech planning and programming (49), impairs verbal WM (52). Given the evidence that WM and LTM are highly interactive (53, 54), auditory LTM for new words may well be preserved in patients with anarthria but not in those with speech apraxia, a predicted dissociation of deficits in need of testing.

Another potential argument against our proposal is the fact that, in infants, recognition of auditory stimuli, including verbal stimuli, develops in advance of expressive language (55, 56), a sequence that would appear to contradict the notion that recognition memory of novel speech sounds requires the participation of a motor circuit. As in anarthria, however, the absence of speech production does not preclude a contribution from the oromotor system to speech recognition memory, and,

indeed, there is evidence (57–59) that even in early infancy speech sounds activate Broca's area. However, perhaps the more important mechanism for "auditory recognition memory" in early infancy is associative memory. Like other nonvocal learners, young infants presumably learn to associate speech sounds with visual and other sensory stimuli, although, in their case, only as a substitute for the more slowly developing and incredibly useful vocal mimicry system.

Importantly, the arcuate fasciculus in humans differs from that in nonhuman primates, including apes. Both its density and complexity increased dramatically during human evolution, presumably driven largely by the highly advantageous appearance and progressive development of speech and language. As described recently in a series of diffusion tensor imaging and resting-state connectivity studies (6, 7, 10, 22, 60), this tract consists not only of a direct connection between caudal superior temporal and ventrolateral frontal areas, but also of a parallel, indirect connection through an intermediate station in the inferior parietal lobule.<sup>†</sup> with this dual-tract system extending ventrally into the middle and inferior temporal gyri as well as dorsally into the ventral premotor gyrus. By comparison, the arcuate fasciculus in monkeys is a primitive one (7, 8, 11, 62), providing a possible explanation for the monkey's apparent inability to store the representations of fluctuating acoustic stimuli in long-term memory (5). Conversely, the large size and complexity of this dual tract in humans may have enabled long-term memory for speech sounds by providing the auditory system with an input that transforms an intractable, fluctuating, acoustic stimulus into an integrated acoustic/oromotor sequence that can be stored for subsequent recognition. If confirmed, the results would imply that speech and auditory memory are so indissolubly linked that neither could have evolved without the other.

## Materials and Methods

**Participants.** Thirty-two young adults (mean age, 29 y; range, 20–40 y; 24 females) took part in the study. Half the participants were native speakers of English, and the other half spoke nonnative English. Twenty-five of the participants had learned to play a musical instrument (mean starting age, 9 y; range, 6–23 y) and continued playing for an average of 7 y.

**Stimuli.** Four types of acoustic stimuli were used: Words, Pseudowords, Nonverbal Sounds, and Reversed Words. The four stimulus types differ in the ease with which they can be reproduced vocally and/or be labeled with a verbal associate and, consequently, in the degree to which speech, language, and musical ability might be able to support their storage for later recognition.

Words, easily reproduced and easily labeled, had a mean duration of 783 msec (range, 600–964) and a mean length of six letters (range, 4–8) and two syllables (range, 1–3).

Pseudowords, easily reproduced but less easily labeled than words, had a mean duration of 756 msec (range, 525–955) and mean length of six letters (range, 4–9) and two syllables (range, 1–3).

<sup>†</sup>This indirect connection observed with diffusion tensor imaging in humans is supported by evidence from tract tracing in monkeys, indicating that the superior temporal gyrus projects to the inferior parietal lobule via the middle longitudinal fasciculus (61), and the inferior parietal lobule projects to the ventral premotor cortex via the superior longitudinal fasciculus (8).

- Mishkin M (1978) Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature* 273:297–298.
- Murray EA, Mishkin M (1984) Severe tactual as well as visual memory deficits follow combined removal of the amygdala and hippocampus in monkeys. *J Neurosci* 4: 2565–2580.
- Murray EA, Mishkin M (1998) Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J Neurosci* 18:6568–6582.
- Buffalo EA, et al. (1999) Dissociation between the effects of damage to perirhinal cortex and area TE. *Learn Mem* 6:572–599.
- Fritz J, Mishkin M, Saunders RC (2005) In search of an auditory engram. *Proc Natl Acad Sci USA* 102:9359–9364.

Nonverbal sounds, difficult to reproduce but labelable, had a duration of 760 msec each, reduced from an original duration of 2–3 s. The auditory stimuli in this category sounded as if they were produced by a xylophone, a boat horn, a saw, bells, etc.

Reversed words, difficult to reproduce or label, were the Words played backward, and so had spectrotemporal compositions and sound envelopes identical to those of the Words.

There were 20 stimuli of each type. In compiling the lists of Words and Pseudowords, we rejected items that were semantically and/or phonologically similar to any of the items we kept. These speech sounds were generated with a speech synthesizer using a UK English female voice (Cepstral) and modified for length and loudness using Adobe Audition 3.0. One hundred Nonverbal Sounds, drawn from the same large sound library that was used by Squire et al. (21) and Fritz et al. (5), were each assigned to one of several categories differing in rhythm, pitch, degree of fluctuation, etc., and then compared by the experimenter with all other sounds in the same category. Only subjectively distinctive stimuli were used to compose the set of 20 Nonverbal Sounds. The intensities of all 80 sounds across the four stimulus types were adjusted by the experimenter to be subjectively equal. For each stimulus type, 10 of the 20 items were randomly selected to form the study list, and the remaining 10 were assigned to be the new items in the recognition test. The procedure for every stimulus type, described below, was adapted from the one used by Squire et al. (21) for nonverbal sounds.

**Procedure. Study list.** Participants listened to a study list containing 10 stimuli of one type and were asked to memorize them. Presentation of the study list took ~12 s (interstimulus intervals, 500 msec), and the presentation order of the 10 stimuli within the list was randomized across participants.

**Filler tasks.** To prevent participants from rehearsing the list between its presentation and the recognition test given 5 min later, they were engaged in one of four different subvocal counting (i.e., articulatory suppression) tasks, two visual and two auditory, taken from the Test of Everyday Attention (TEA) (63). The visual tasks required counting of symbols (either on a city map or in a simulated telephone directory), and the auditory tasks required counting of tones (either all tones if only low tones were presented or only low tones if low and high tones were presented). Each task filled the 5-min delay interval, including explanation of the task by the experimenter (~2 min) and test time (~3 min). Each participant received a different filler task after each study list, and the four filler tasks were counterbalanced across participants and stimulus types.

**Recognition test.** After the filler task, participants listened to 20 stimuli (the 10 from the study list and the 10 novel ones) presented in a randomized order, and, after each stimulus, indicated whether it was "old" or "new." They then pressed a keyboard space bar to hear the next sound. The same procedure—study list, filler task, recognition test—was repeated for each type of stimulus, with the order of stimulus type counterbalanced across participants using a Latin square design.

**Discrimination task.** After completing the recognition memory test, participants were tested for their ability to discriminate reversed words. In this task, the participants listened to 40 pairs of reversed words (20 "same" pairs and 20 "different" pairs) and indicated whether the two members of each pair were the same or different (intrapair interval, 500 msec; mean difference in duration of intrapair stimuli, 44 msec, range 2–93 msec). The response to each pair was self-paced, and participants pressed the space bar to hear the next pair. Four different orders of the 40 stimulus pairs were programmed for presentation, and each order was administered to eight randomly selected participants.

**ACKNOWLEDGMENTS.** This study was supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, as well as by the Medical Research Council (G0300117/65439) and University College London Pump Prime Grant 06CN05.

- Catani M, et al. (2007) Symmetries in human brain language pathways correlate with verbal recall. *Proc Natl Acad Sci USA* 104:17163–17168.
- Rilling JK, et al. (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat Neurosci* 11:426–428.
- Petrides M, Pandya DN (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol* 7:e1000170.
- Saur D, et al. (2008) Ventral and dorsal pathways for language. *Proc Natl Acad Sci USA* 105:18035–18040.
- Frey S, Campbell JS, Pike GB, Petrides M (2008) Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci* 28: 11435–11444.

11. Rilling JK, Glasser MF, Jbabdi S, Andersson J, Preuss TM (2012) Continuity, divergence, and the evolution of brain language pathways. *Front Evol Neurosci* 3:1–6.
12. Samson S (1999) Musical function and temporal lobe structures: A review of brain lesion studies. *J New Music Res* 28:217–228.
13. Samson S, Zatorre RJ (1992) Learning and retention of melodic and verbal information after unilateral temporal lobectomy. *Neuropsychologia* 30:815–826.
14. DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci USA* 109:E505–E514.
15. Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Front Syst Neurosci* 5:1.
16. Gardiner JM, Brandt KR, Vargha-Khadem F, Baddeley AD, Mishkin M (2006) Effects of level of processing but not of task enactment on recognition memory in a case of developmental amnesia. *Cogn Neuropsychol* 23:930–948.
17. Brandt KR, Gardiner JM, Vargha-Khadem F, Baddeley AD, Mishkin M (2008) Impairment of recollection but not familiarity in a case of developmental amnesia. *Neurocase* 15:60–65.
18. Sagar HJ, Gabrieli JD, Sullivan EV, Corkin S (1990) Recency and frequency discrimination in the amnesic patient H.M. *Brain* 113:581–602.
19. Haist F, Shimamura AP, Squire LR (1992) On the relationship between recall and recognition memory. *J Exp Psychol Learn Mem Cogn* 18:691–702.
20. Zola-Morgan S, Squire LR, Amaral DG (1986) Human amnesia and the medial temporal region: Enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *J Neurosci* 6:2950–2967.
21. Squire LR, Schmolck H, Stark SM (2001) Impaired auditory recognition memory in amnesic patients with medial temporal lobe lesions. *Learn Mem* 8:252–256.
22. Catani M, Jones DK, ffytche DH (2005) Perisylvian language networks of the human brain. *Ann Neurol* 57:8–16.
23. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461.
24. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36.
25. di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: A neurophysiological study. *Exp Brain Res* 91:176–180.
26. Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res* 3:131–141.
27. Kohler E, et al. (2002) Hearing sounds, understanding actions: Action representation in mirror neurons. *Science* 297:846–848.
28. Rizzolatti G, Craighero L (2004) The mirror-neuron system. *Annu Rev Neurosci* 27:169–192.
29. Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychon Bull Rev* 13:361–377.
30. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702.
31. Pulvermüller F, Hauk O, Nikulin VV, Ilmoniemi RJ (2005) Functional links between motor and language systems. *Eur J Neurosci* 21:793–797.
32. Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage* 33:316–325.
33. Meister IG, Wilson SM, Debleick C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. *Curr Biol* 17:1692–1696.
34. Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci* 29:9819–9825.
35. Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur J Neurosci* 15:399–402.
36. Lotto AJ, Hickok GS, Holt LL (2009) Reflections on mirror neurons and speech perception. *Trends Cogn Sci* 13:110–114.
37. Hickok G (2009) Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci* 21:1229–1243.
38. Hickok G (2009) The functional neuroanatomy of language. *Phys Life Rev* 6:121–143.
39. Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69:407–422.
40. Hickok G, Costanzo M, Capasso R, Miceli G (2011) The role of Broca's area in speech perception: Evidence from aphasia revisited. *Brain Lang* 119:214–220.
41. Yeterian EH, Pandya DN (1991) Prefrontostriatal connections in relation to cortical architectonic organization in rhesus monkeys. *J Comp Neurol* 312:43–67.
42. Price CJ, Crinion JT, MacSweeney M (2011) A generative model of speech production in Broca's and Wernicke's areas. *Front Psychol* 2:237.
43. Seltzer B, Pandya DN (1989) Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 281:97–113.
44. Romanski LM, et al. (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136.
45. Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–310.
46. Mooney R (2009) Neural mechanisms for learned birdsong. *Learn Mem* 16:655–669.
47. Lenneberg EH (1962) Understanding language without ability to speak: A case report. *J Abnorm Soc Psychol* 65:419–425.
48. Christen HJ, et al. (2000) Foix-Chavany-Marie (anterior operculum) syndrome in childhood: A reappraisal of Worster-Drought syndrome. *Dev Med Child Neurol* 42:122–132.
49. Liégeois FJ, Morgan AT (2012) Neural bases of childhood speech disorders: Lateralization and plasticity for speech functions during development. *Neurosci Biobehav Rev* 36:439–458.
50. Bishop DVM, Robson J (1989) Unimpaired short-term-memory and rhyme judgment in congenitally speechless individuals: Implications for the notion of articulatory coding. *Q J Exp Psychol* 41:123–140.
51. Baddeley AD, Wilson B (1985) Phonological coding and short-term-memory in patients without speech. *J Mem Lang* 24:490–502.
52. Waters GS, Rochon E, Caplan D (1992) The role of high-level speech planning in rehearsal: Evidence from patients with apraxia of speech. *J Mem Lang* 31:54–73.
53. Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol Bull* 104:163–191.
54. Cowan N (2008) What are the differences between long-term, short-term, and working memory? *Prog Brain Res* 169:323–338.
55. Trainor LJ, Wu L, Tsang CD (2004) Long-term memory for music: Infants remember tempo and timbre. *Dev Sci* 7:289–296.
56. Spence MJ (1996) Young infants' long-term auditory memory: Evidence for changes in preference as a function of delay. *Dev Psychobiol* 29:685–695.
57. Imada T, et al. (2006) Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *Neuroreport* 17:957–962.
58. Dehaene-Lambertz G, et al. (2006) Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc Natl Acad Sci USA* 103:14240–14245.
59. Bristow D, et al. (2009) Hearing faces: How the infant brain matches the face it sees with the speech it hears. *J Cogn Neurosci* 21:905–921.
60. Kelly C, et al. (2010) Broca's region: Linking human brain functional connectivity data and non-human primate tracing anatomy studies. *Eur J Neurosci* 32:383–398.
61. Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1–24.
62. Thiebaut de Schotten M, Dell'Acqua F, Valabregue R, Catani M (2012) Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex* 48:82–96.
63. Robertson IH, Ward T, Ridgeway V, Nimmo-Smith I (2001) The Test of Everyday Attention (TEA). *Test Reviews* 4:51–55.