More information is more information: Working memory and visual context

Joanna H. Lowenstein and Susan Nittrouer
Department of Speech, Language, and Hearing Sciences, University of Florida

BACKGROUND

Listeners with cochlear implants demonstrate diminished auditory-verbal working memory capacities, possibly due to a lack of durable codes in the memory buffer (Nittrouer et al, 2017; Moberly et al, 2017). Earlier studies which examined the suffix effect suggest that lip-read information should enhance those codes (Campbell & Dodd, 1980; Spoehr & Corin, 1978). According to these earlier studies, both the phonological structure and the dynamic nature of lip-read signals contribute to this facilitative effect.

The dual-component memory model used in this study consists of a front end that recovers some form of structure from the sensory input and deposits that structure into a temporary store, and a central processor that performs operations on that recovered and temporarily stored structure (Baddeley & Hitch, 1974). Figure 1 presents a visualization of the model.

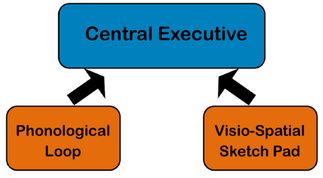

Figure 1. Working memory model by Baddeley & Hitch (1974). Language materials are processed through a phonological loop that extracts phonological structure. Visual signals are processed through a sketch pad, which recovers spatial structure.

PRIMARY HYPOTHESIS

The primary hypothesis tested in this study was that lip-read information should uniquely enhance the operation of the phonological loop for degraded speech signals.

To test this hypothesis, three kinds of signals were used:

- Unprocessed speech signals consisting of clear phonological structure
- Vocoded speech signals, consisting of degraded phonological structure
- Nonspeech environmental sounds that were clear, but did not consist of phonological structure

Two kinds of visual signals were added:

- Dynamic signals specifying the event that generated the signal
- Pictures representing the object named or creating the signal

Predictions:
Clear speech would provide optimally durable codes, so no enhancement would be observed with either visual signal.

Vocoded speech would benefit from lip-read signals only, because these signals are phonological in structure and dynamic.

Environmental sounds would benefit from neither kind of visual signal, because they are not phonological in structure; pictures should be processed in the visio-spatial sketch pad, so provide no benefits.

METHODS

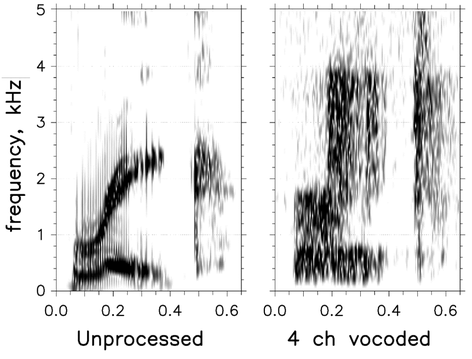

Figure 2. Spectrogram of "Rake," unprocessed (left) and 4-channel vocoded (right)

Eighty young adult listeners with normal hearing were asked to recall order of 3 kinds of auditorily presented stimuli (8 items) in a closed-set format: unprocessed monosyllabic nouns (UP); those same nouns, only noise vocoded (VOC); and environmental sounds (Sounds).

Those listeners also recalled order in one of two audio-visual conditions (40 listeners per condition):

1) Observing the movement that created the noun or sound
2) Seeing a picture of the object being labeled (nouns) or creating the sound (environmental sounds).

Dependent measures were:

1) Accuracy of recalling the order of presentation
2) Response time, which served to index cognitive load

Figure 3. Participant completing the serial recall task

RESULTS

ACCURACY

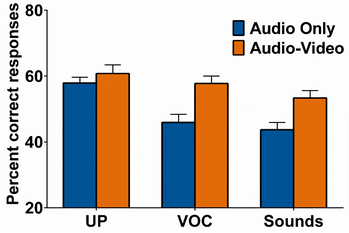

Figure 4. Percent correct responses for the dynamic (Audio-Visual) condition. Participants demonstrated the greatest gain in accuracy for the VOC stimuli, but they also demonstrated greater accuracy for Sounds. The change in accuracy for UP was not significant.

	UP	VOC	Sounds
t	1.22	6.17	5.94
p	NS	<.001	<.001
d	.21	.80	.70

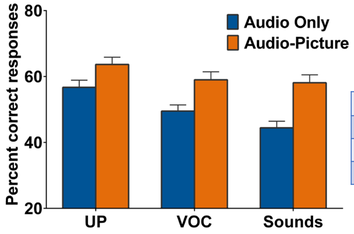

Figure 5. Percent correct responses for the static (Audio-Picture) condition. Participants demonstrated gains in accuracy for all three sets of stimuli.

	UP	VOC	Sounds
t	3.92	3.88	7.01
p	<.001	<.001	<.001
d	.51	.71	1.01

TIME

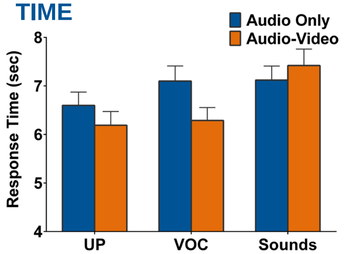

Figure 6. Mean task time for the Audio-Video condition. Participants were faster for the UP and VOC stimuli, but slower for environmental sounds. The largest improvement in task time was for the UP stimuli.

	UP	VOC	Sounds
t	-2.44	-3.26	2.06
p	.02	.002	.046
d	-.24	-.45	.15

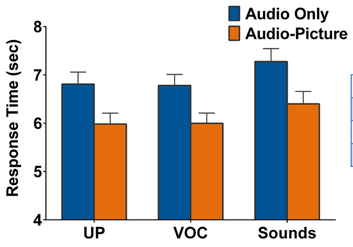

Figure 7. Mean task time for the Audio-Picture condition. Participants were faster for all three conditions.

	UP	VOC	Sounds
t	-4.01	-3.97	-5.44
p	<.001	<.001	<.001
d	-.60	-.56	-.54

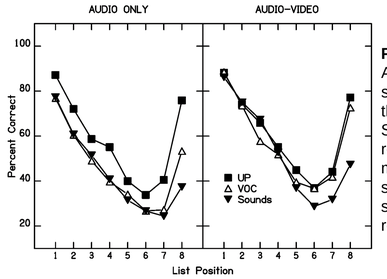

Figure 8. Position accuracy for the Audio-Video condition, for all three sets of stimuli. The VOC stimuli in the Audio-Only looked more like Sounds, with reduced primacy and recency effects. Adding dynamic lip movements resulted in the VOC stimuli being processed as UP stimuli, with the same primacy and recency effects.

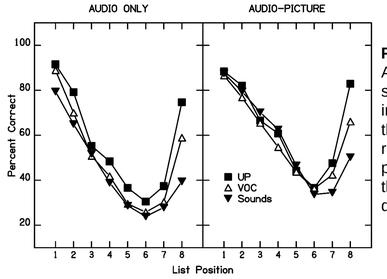

Figure 9. Position accuracy for the Audio-Picture condition, for all three sets of stimuli. Adding static pictures improved performance across all three sets of stimuli, but did not result in VOC stimuli being processed the same as UP stimuli—the recency effect was still diminished.

CONCLUSION

Adding lip movements uniquely enhances the operation of the phonological loop for degraded speech signals, improving overall accuracy and the effects of both primacy and recency.

Adding static pictures did improve overall accuracy for degraded speech signals, but did not improve recency in the same way as lip movements. Static pictures did improve overall accuracy more than dynamic movements for the unprocessed speech and sound stimuli; they also resulted in reduced cognitive load (faster task times) across all three sets of stimuli.

REFERENCES

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-90). Academic Press: San Diego, CA.

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Q. J. Exp. Psychol.*, *32*, 85-99.

Moberly, A. C., Harris, M. S., Boyce, L., & Nittrouer, S. (2017). Speech recognition in adults with cochlear implants: The effects of working memory, phonological sensitivity, and aging. *J. Speech Lang. Hear. Res.*, *60*, 1046-1061.

Nittrouer, S., Caldwell-Tarr, A., Low, K. E., & Lowenstein, J. H. (2017). Verbal working memory in children with cochlear implants. *J. Speech Lang. Hear. Res.*, *60*, 3342-3364.

Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Mem. Cognit.*, *6*, 583-589.

ACKNOWLEDGEMENTS

Work supported by NIDCD Grant No DC 000633. The authors thank Shay Fitzgerald and Yi Yuan for their help with data collection.