

On advantage of seeing TEXT and hearing SPEECH

Jelena Živanović¹ and Dušica Filipović Đurđević^{1&2}

¹ *Department of Psychology, Faculty of Philosophy, University of Novi Sad, Serbia*

² *Laboratory for Experimental Psychology, Faculty of Philosophy,
University of Belgrade, Serbia*

The aim of this study was to examine the effect of congruence between the sensory modality through which a concept can be experienced and the modality through which the word denoting that concept is perceived during word recognition. Words denoting concepts that can be experienced visually (e.g. “color”) and words denoting concepts that can be experienced auditorily (e.g. “noise”) were presented both visually and auditorily. We observed shorter processing latencies when there was a match between the modality through which a concept could be experienced and the modality through which a word denoting that concept was presented. In visual lexical decision task, “color” was recognized faster than “noise”, whereas in auditory lexical decision task, “noise” was recognized faster than “color”. The obtained pattern of results can not be accounted for by exclusive amodal theories, whereas it can be easily integrated in theories based on perceptual representations.

Key words: sensory modality, conceptual representations, Perceptual Symbol Theory

We experience the world around us by using our senses. Most of the objects surrounding us can be experienced through several sensory modalities. For example, *rooster* can be experienced through at least five sensory modalities – it can be seen, heard, touched, smelled, and even tasted. Also, there are objects that can be experienced through only one of the sensory modalities. Some objects can only be experienced visually (e.g. *rainbow, sky, moon*), while others can only be experienced auditorily (e.g. *chirping, thunder, noise*). At the same time, words denoting those objects can be presented visually, as in reading, or auditorily, as in verbal communication (or even through tactile modality, as in Braille alphabet).

The aim of this study was to examine whether processing time is affected by the congruence between the sensory modality through which a concept is

Corresponding author: dfdurdevic@ff.uns.ac.rs

* This work was supported by the Ministry of Science and Technological Development of Serbia, grant number 179006.

experienced and the modality through which the word denoting that concept is perceived during word recognition. The answer to this question is highly dependent on the theoretical viewpoint. According to amodal theories of conceptual representations (Fodor, 1975; Pylyshyn, 1984; Smith & Medin, 1981; Tulving, 1972), sensorimotor representations are transduced into amodal representations, such as feature lists or semantic networks. These amodal representations have no correspondence to the perceptual states that produced them and are arbitrarily linked to them. Amodal symbols that represent concepts in their absence reside in a different neural system from the representations of these concepts during perception itself. In addition, these two systems operate according to different principles. Therefore, according to amodal theories, the congruence between the sensory modality through which a concept is experienced and the modality through which the word denoting that concept is perceived should not affect word recognition. On the other hand, there are theories, such as Perceptual Symbol Theory (Barsalou, 1999), which postulate that our knowledge is grounded in a modality-specific system. According to this theory, concept representations (named perceptual symbols) are based on perceptual experience related to the concept and can be seen as records of neural states that underlie perception. Therefore, the activation of the concept representation would involve simulation of the specific experience, that is re-enactment of neural activation patterns that were active in the experience with that concept. Perceptual symbols reside in the same system as perceptual states that produce them. Each type of a symbol becomes established in its respective brain area – visual symbols become established in visual areas, auditory symbols in auditory areas, proprioceptive symbols in somatosensory and motor areas, and so forth. This claim is supported by findings from cognitive neuroscience, as well as behavioral studies (Barsalou, 1999; Pulvermüller, 1999; Šetić & Domijan, 2007). Therefore, based on the assumptions of Perceptual Symbol Theory, it could be derived that congruence between the sensory modality through which a concept is experienced and the modality through which the word denoting that concept is perceived should affect word recognition.

In this study, only words denoting concepts that can be experienced through one sensory modality were considered. Two groups of words were presented – words denoting concepts that can be experienced only through visual modality (e.g. *rainbow*, *sky*, *moon*) and words denoting concepts that can be experienced only through auditory modality (e.g. *chirping*, *thunder*, *noise*). The words denoting two groups of concepts were presented visually (in visual lexical decision task) and auditorily (in auditory lexical decision task). Based on the Perceptual Symbol Theory, it was hypothesized that words denoting concepts that can be visually experienced would be faster and more accurately recognized in visual lexical decision task than words denoting concepts that can be auditorily experienced. On the other hand, in auditory lexical decision task, words denoting concepts that can be auditorily experienced would be faster and more accurately recognized than words denoting concepts that can be visually experienced.

Method

Participants: Thirty-nine undergraduate students from the Department of Psychology, Faculty of Philosophy, University in Novi Sad were randomly assigned to one of the two experimental blocks (either visual lexical decision task or auditory lexical decision task). All were native speakers of Serbian and had normal hearing and normal or corrected-to-normal vision.

Materials and design: A total of sixty Serbian nouns (in nominative singular) and 60 pseudo-nouns were presented in two lexical decision tasks.

Critical stimuli were two groups of nouns – 20 nouns denoting concepts that can be experienced only through visual modality (e.g. *sky, rainbow, moon*) and 20 nouns denoting concepts that can be experienced only through auditory modality (e.g. *chirping, thunder, noise*). The two groups were considered as two levels (auditory/visual) of the factor that was named “concept modality” for the purposes of the current experiment. Both groups were presented in two lexical decision tasks – visual lexical decision task and auditory lexical decision task. This way, concept modality and presentation modality (task) were crossed in a 2x2 factorial design.

The selection of the two groups of nouns was based on ratings of concreteness-by-separate-modalities, which included separate ratings for visual, auditory, gustative, olfactory and tactile modality (Popović, Živanović & Filipović Đurđević, 2009). The selection criterion included only nouns that were high on either visual or auditory modality, and low on all remaining modalities. The two groups were matched for printed frequency (Kostić, 1999), word familiarity (subjective frequency), word length in letters/phonemes/syllables, Coltheart’s N (Coltheart, Davelaar, Jonasson, & Besner, 1977) and uniqueness point (Radeau, Mousty, & Bertelson, 1989; Turner, Valentine, & Ellis, 1998). Despite the effort to match the two critical groups of words for general concreteness of the denoted concepts, there remained significant difference between the two groups in the general concreteness ratings: $F(1, 38) = 12.83, p < 0.001$. The group of words denoting concepts that can be experienced only through auditory modality was higher in general concreteness ($M = 5.13$) than the group of words denoting concepts that can be experienced only through visual modality ($M = 4.73$). Therefore, the variable “general word concreteness” was controlled in the analyses, in the way of taking it as co-variable. Critical stimuli and their relevant characteristics are listed in the Appendix.

Along with two critical groups of nouns, we presented participants with the group of fillers consisting of 20 nouns denoting concepts that can be experienced through several modalities. Words from different categories were used (food, animals, objects...) in order to prevent participants of encountering only natural phenomena (the most common in the group with visual modality) and sounds (in the auditory group) as stimuli. Although they were not analyzed, fillers were matched with the two critical groups of words for frequency and length. Finally, 60 pseudo-nouns were derived from the novel set of nouns. Nouns and pseudo-nouns were matched for length.

For purposes of auditory lexical decision task, the stimuli were recorded by a computer and specialized software “Praat” (Boersma & Weenink, 2009). During the recording, in accordance with previous studies (Słowiaczek & Pisoni, 1986), adult male person pronounced stimuli in the sentence “Say *stimulus* please.” In the next step, using the same software, words were extracted from the carrying sentence. All groups of words and pseudo-words were matched for pronunciation duration (in milliseconds).

Dependant variables were reaction time measured from the moment of stimulus onset (in milliseconds) and error probability.

Procedure: The stimuli were presented using *SuperLab Pro 2.0* (Cedrus, 2001). In visual lexical decision task stimuli were presented visually, on the screen, whereas in auditory lexical decision task, stimuli were presented auditorily, binaurally, by headphones. Each trial was preceded with a fixation point at the centre of the screen in duration of 1500 ms. Maximal duration of stimuli presentation was limited to 1500ms. Responses were given by button-press. Reaction times were measured from the presentation of the stimulus until the button press. Prior to experiment, 12 practice trials were presented that were not analyzed. The order of stimuli presentation was randomized across participants.

Results

Prior to analysis, data obtained from one participant from auditory lexical decision task were excluded, due to large number of errors. Additionally, items that elicited above 20% of errors were excluded from analyses of response latencies (six words from auditory lexical decision task and four words from visual lexical decision task).

Response latencies: In by-participant analysis, we performed 2x2 analysis of variance. Because the two critical groups of nouns were not perfectly matched for general concreteness, this variable was statistically controlled in by-item analysis of covariance. There was a significant effect of task ($F(1, 36)=21.768$, $p<0.0001$, by participant) and a significant interaction of task and concept modality ($F(1, 36)=15.609$, $p<0.001$, by participant; $F(1, 30)=7.669$, $p<0.01$, by item). Our participants were generally faster in visual lexical decision task. However, the effect of concept modality was dependent on the task performed. In visual lexical decision task, words denoting concepts that can be experienced through visual modality elicited shorter response latencies when compared to words denoting concepts that can be experienced through auditory modality ($t(17)=-2.8$, $p<0.05$; Tukey HSD: $p<0.05$, by participant). On the other hand, when presented in auditory lexical decision task, words denoting concepts that can be experienced through visual modality elicited longer response latencies when compared to words denoting concepts that can be experienced through auditory modality ($t(19)=2.777$, $p<0.05$; Tukey HSD: $p<0.05$, by participant). Although post-hoc tests did not reach significance in by-item analysis, the general pattern of results was similar to that obtained in by-participant analysis.

Additionally, in by-item analysis, we applied general linear model in which modality as a discrete variable was replaced with continuous ratings of visibility/audibility of the concept, that is with ratings of the extent of possibility to experience a given concept by visual/auditory modality. Because due to the nature of the stimuli selection process, the two measures were highly correlated ($r=-0.95$), we performed two separate analyses, one for each of the two measures. In both analyses, in step-wise manner, we controlled for the effects of task and general concreteness prior to including audibility/visibility ratings in the model. Additionally, we looked at all possible interactions.

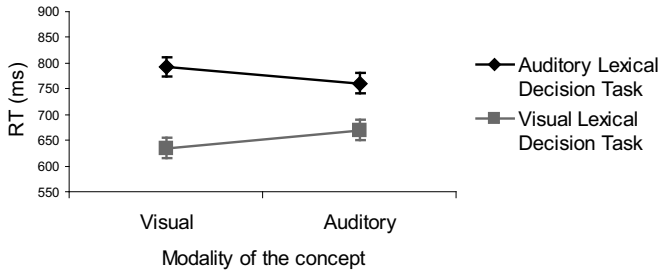


Figure 1. The interaction of modality through which the concept can be experienced and task, that is modality through which the word denoting given concept is presented (vertical bars denote Standard Error of the Mean).

In the first analysis, along with significant main effect of task ($F(1, 30)=108.509, p<0.01$), there was a significant interaction of task and concept visibility ($F(1, 30)=7.018, p<0.05$). A more detailed look into the observed interaction revealed that the effect of visibility ratings was present only at the level of visual lexical decision task, facilitating processing of visually presented words ($\beta=-0.455, t(30)=-2.55, p<0.05$). Importantly, there was no interaction of task and general concreteness. The effect of general concreteness was marginal ($F(1, 30)=2.651, p=0.114$) and facilitative in both tasks.

In the second analysis, along with significant main effect of task ($F(1, 30)=124.553, p<0.01$), there was a significant interaction of task and concept audibility ($F(1, 30)=12.491, p<0.01$). As in the previous analysis, the effect of audibility ratings was limited to visual lexical decision task, where it inhibited processing of visually presented words ($\beta=0.531, t(30)=2.68, p<0.05$). Although there was a trend towards facilitative effect of audibility ratings on auditorily presented words, this effect did not reach significance ($\beta=-0.194$). Importantly, as in the previous analysis, there was no interaction of task and general concreteness.

Error counts: Error analysis was performed by applying logistic regression model to binomial distribution of correct responses and errors. We looked at main effects of task and modality and tested for their interaction. Because the two critical groups of nouns were not perfectly matched for general concreteness, this variable was included in the model, as a control variable. The analysis revealed that there was a significant main effect of general concreteness ($\chi^2_{(1)}=6.72, p<0.01$), significant main effect of task ($\chi^2_{(1)}=6.45, p<0.05$) and of modality ($\chi^2_{(1)}=4.02, p<0.05$). However, there was no interaction. Probability of error was negatively correlated with general concreteness ($\beta=-1.156, z=-2.59, p<0.01$). At the same time, error counts were lower in visual lexical decision task ($\beta=-0.829, z=-2.54, p<0.05$) and for nouns that can be experienced through visual modality ($\beta=-0.805, z=-2.01, p<0.05$).

DISCUSSION AND CONCLUSION

In this study, we hypothesized that word recognition would be affected by congruence between the sensory modality through which a concept is experienced and the modality through which the word denoting that concept is perceived. As predicted, in visual lexical decision task, we observed that words denoting concepts that can be experienced only through visual modality were recognized faster than words denoting concepts that can be experienced only through auditory modality. Along the same line, in auditory lexical decision task, words denoting concepts that can be experienced only through auditory modality were recognized faster than words denoting concepts that can be experienced only through visual modality.

The obtained pattern of results can not be accounted for by amodal theories (Fodor, 1975; Pylyshyn, 1984; Smith & Medin, 1981; Tulving, 1972). According to these theories, amodal symbols that represent concepts in the absence of perception are grounded in the system which is separate from perceptual. Symbolic concept representations have no similarities to perceptual characteristics of the represented concept. Although the information about the modality through which the concept is experienced can be recorded as one of the concept's characteristics, this information is established in the different form. Therefore, the effect of presentation modality should not depend on modality-specific characteristics of the concept, which was not the case in this study.

The obtained pattern of results can easily be accounted for by Perceptual Symbol Theory, and is in accordance with numerous findings that bring evidence in favor of existence of sensory-modality-specific system of knowledge conceptualization (Barsalou, 1999; Estes, Verges, & Barsalou, 2008; Pulvermüller, 1999; Stenberg, Radeborg, & Hedman, 1995; Šetić & Domijan, 2007). Perceptual symbols are hypothesized to be grounded in the same system in which perceptual experiences they are based on are, in suitable brain areas, depending on the modality (visual symbols are established in visual areas, auditory symbols in auditory areas, proprioceptive symbols in somatosensory and motor areas, and so forth [Barsalou, 1999]). Therefore, auditorily presented words would activate auditory brain areas, leading to fast activation of auditory perceptual symbols, because their re-enactment is based on the same mechanisms as perception of auditory stimuli. This would account for the advantage of words denoting concepts with auditory modality relative to words denoting concepts with visual modality when stimuli are presented auditorily. On the other hand, visual presentation of words would highly activate visual areas enabling faster recognition of words denoting concepts with visual modality, symbols of which are established in these areas.

Although our findings can be integrated in Perceptual Symbol Theory, while presenting a problem for amodal theories, they do not bring any further

tests of a more detailed explanation within the framework of the given theory. The observed interaction could be based on two different processes. On the one hand, there could be some facilitation in the congruent conditions compared to incongruent conditions, that is the activation of a modality usually activated when experiencing an object could help processing. On the other hand, there could be some inhibition in the incongruent conditions compared to congruent conditions, that is the activation of a modality that is not activated when experiencing an object could make processing more demanding. Only the first possibility is implied in the interpretation we provided. The presented experiment was not designed to test for this difference, and future research should address this issue. One way to do so would be to apply a more sophisticated correlation design, that is to perform some regression modeling. The regression analysis applied in this paper did not bring unambiguous answer as this experiment was not designed for this purpose. In addition to facilitation by congruency or inhibition by incongruence dilemma, there remains a question of the nature of activation that facilitates or inhibits processing. In this experiment, we believe that there was some kind of unspecific, general activation present. However, a question remains of what would happen if participants were engaged in some kind of more specific processing (for example, see Estes, Verges, & Barsalou, 2008 for a reverse pattern of results in a more specific task).

The hypothesized interaction that was observed in response latencies was not observed in error counts. Probability of error was lower for words denoting concepts that can be experienced through visual modality, regardless of the presentation modality. The precise reason of the observed advantage of visual modality in recognition accuracy remains as an open question and a challenge for future studies. One way to disentangle this inconsistency would be to engage participants in different experimental tasks.

REFERENCES

- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009 from <http://www.praat.org/>
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*, (pp. 535–555). Hillsdale, NJ: Erlbaum.
- Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head Up, Foot Down. *Psychological science*, 19(2), 93–97.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Kostić, Đ. (1999). *Frekvencijski rečnik savremenog srpskog jezika*. Tom I – VII. Institut za eksperimentalnu fonetiku i patologiju govora, Beograd i Laboratorija za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu.
- Popović, M., Živanović, J. & Filipović Đurđević, D. (2009). Uticaj broja i vrste čulnih modaliteta na procenu konkretnosti reči i brzinu obrade reči. *XV naučni skup Empirijska istraživanja u psihologiji*. Beograd, *Rezimea*, pp. 24–25.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253–336.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, 51(3), 123–128.
- Slowiaczek, L. M. & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, 14(3), 230–237.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Stenberg, G., Radeborg, K., & Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. *Memory & Cognition*, 23(4), 425–441.
- Šetić, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, 22(2), 297–312.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*, (pp. 381–403). New York: Academic Press.
- Turner, J., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition*, 26(6), 1282–1291.

APPENDIX

Word	Word length (in letters)	Lemma frequency	Concreteness	Familiarity	Uniqueness point	Coltheart's N	Pronunciation duration (ms)	RT in Visual Lexical Decision Task (ms)	RT in Auditory Lexical Decision Task (ms)	Concept modality
bljesak	6	133	4.796	6.385	2	1	491	660	765	Visual
boja	4	830	5.225	7	4	18	260	532	768	Visual
duga	4	180	4.552	6.923	4	13	295	600	700	Visual
horizont	8	99	3.959	6.923	8	0	558	651	817	Visual
internet	8	0	3.938	6.769	7	2	541	645	813	Visual
iskra	5	136	4.449	6.308	5	0	330	676	818	Visual
kometa	6	11	4.184	6.308	5	2	468	708	764	Visual
magla	5	631	4.776	6.692	5	2	382	572	749	Visual
mesečina	8	424	4.612	6.923	6	0	594	627	900	Visual
mimika	6	5	4.918	6.769	4	0	418	741	820	Visual
pomračina	9	1	4.816	6.615	8	0	550	707	892	Visual
munja	4	216	4.531	6.692	4	8	348	631	724	Visual
nebo	4	2818	4.776	7	4	4	363	560	715	Visual
oblak	5	1565	5.041	7	5	2	328	589	656	Visual
odsajaj	6	67	4.735	6.769	4	1	330	750	914	Visual
osmeh	5	720	5.408	7	5	1	438	563	755	Visual
senka	5	1209	4.265	6.769	4	5	457	608	839	Visual
slovo	5	248	5	6.923	5	1	415	607	830	Visual
tekst	5	87	5.489	7	5	1	356	593	817	Visual
varnica	7	72	5.163	6.385	6	3	507	738	762	Visual
grmljavina	9	40	5.184	6.692	5	0	561	710	954	Auditory
grom	4	115	5.02	6.769	4	11	267	643	708	Auditory
pucketanje	9	3	4.898	6.692	8	0	608	791	869	Auditory
žubor	5	76	4.878	6.077	3	0	381	722	829	Auditory
buka	4	60	4.959	6.769	4	25	343	622	542	Auditory
cvrkut	6	86	5.102	6.769	4	0	510	710	805	Auditory

galama	6	16	5	6.692	5	2	462	637	680	Auditory
glas	4	1513	5.449	7	4	5	382	633	739	Auditory
melodija	8	78	5.327	6.846	7	1	455	647	762	Auditory
muzika	6	446	5.694	7	6	0	470	555	679	Auditory
odjek	5	164	4.938	6.462	5	2	370	654	857	Auditory
pucanj	5	95	5.327	6.615	5	2	422	650	822	Auditory
šapat	5	302	4.979	6.923	5	3	479	740	823	Auditory
škripa	6	45	4.939	6.462	6	1	439	802	892	Auditory
ton	3	64	4.939	6.846	3	14	234	629	784	Auditory
zveket	6	32	5.041	6	5	0	491	794	876	Auditory
zvuk	4	656	5.082	7	4	2	346	580	686	Auditory
jauk	4	125	5.082	6.385	3	5	410	771	707	Auditory
mjauk	5	0	5.102	6.538	3	0	424	687	788	Auditory
vrisak	6	140	5.633	6.615	4	2	462	633	651	Auditory
