

Linfield College DigitalCommons@Linfield

Senior Theses

Student Scholarship & Creative Works

2012

The Influence of Social Priming on Speech Perception

Sierra N. Broussard Linfield College

Follow this and additional works at: https://digitalcommons.linfield.edu/psycstud_theses

Part of the Psychology Commons

Recommended Citation

Broussard, Sierra N., "The Influence of Social Priming on Speech Perception" (2012). *Senior Theses*. 2. https://digitalcommons.linfield.edu/psycstud_theses/2

This Thesis (Open Access) is protected by copyright and/or related rights. It is brought to you for free via open access, courtesy of DigitalCommons@Linfield, with permission from the rights-holder(s). Your use of this Thesis (Open Access) must comply with the Terms of Use for material posted in DigitalCommons@Linfield, or with other stated terms (such as a Creative Commons license) indicated in the record and/or on the work itself. For more information, or if you have questions about permitted uses, please contact digitalcommons@linfield.edu.

Running head: SOCIAL PRIMING AND SPEECH PERCEPTION

LINFIELD COLLEGE

The Influence of Social Priming on Speech Perception

A thesis submitted in partial satisfaction of the requirements for

Departmental Honors in Psychology

By

Sierra Noel Broussard

2012

=

The thesis of Sierra Broussard is approved:

Signature redacted Lee Bakner, PhD

Signature redacted Eugene Gilden, PhD

Signature redacted

Chuck Dunn, PhD

Signature redacted

Kay Livesay, PhD Thesis Advisor

May 21, 2012

Ξ

LINFIELD COLLEGE CHRP-IRB Chair (503) 883-2708

MEMORANDUM

DATE: February 14, 2012

TO: Sierra Broussard

FROM: **Kay Livesay** Committee for Human Research Participation (CHRP) Institutional Review Board (IRB) IORG0002606

RE: CHRP/IRB 201112-14 Title of research: Language Comprehension

The IRB has reviewed the above titled application and grant IRB approval to the research. You may start collecting data. As the primary investigator for this project you hereby accept the responsibility to comply with the standards and requirements established by the Belmont Report to protect the rights and welfare of human subjects involved in research conducted under this application. By preceding with your research you acknowledge that you are responsible for protecting the rights and welfare of each research subject, and that the subject's rights and welfare must take precedence over the goals of the research.

If any changes to your protocol are necessary, they must be reviewed by the IRB prior to the institution of the changes in your research. Please be aware that you must use the informed consent I have sent you because it contains the IRB approval stamp.

If you have any questions or concerns, please contact me at (503) 883-2708.

Thank you,

Signature redacted

Kay Livesay Associate Professor of Psychology Chair IRB klivesa@linfield.edu (503) 883-2708

THESIS COPYRIGHT PERMISSIONS

Please read this document carefully before signing. If you have questions about any of these permissions, please contact the <u>DigitalCommons Coordinator</u>.

Title of the Thesis:

The Influence of Social Priming on Speech Perception

Author's Name: (Last name, first name)

Broussard, Sierra Noel

Advisor's Name

Kay Livesay

DigitalCommons@Linfield is our web-based, open access-compliant institutional repository for digital content produced by Linfield faculty, students, staff, and their collaborators. It is a permanent archive. By placing your thesis in DigitalCommons@Linfield, it will be discoverable via Google Scholar and other search engines. Materials that are located in DigitalCommons@Linfield are freely accessible to the world; however, your copyright protects against unauthorized use of the content. Although you have certain rights and privileges with your copyright, there are also responsibilities. Please review the following statements and identify that you have read them by signing below. Some departments may choose to protect the work of their students because of continuing research. In these cases, the project is still posted in the repository but content will only be accessible by individuals who are part of the Linfield community.

CHOOSE THE STATEMENT BELOW THAT DEFINES HOW YOU WANT TO SHARE YOUR THESIS. THE FIRST STATEMENT PROVIDES THE MOST ACCESS TO YOUR WORK; THE LAST STATEMENT PROVIDES THE LEAST ACCESS.

<u>×</u> I **agree** to make my thesis available to the Linfield College community <u>and</u> to the larger scholarly community upon its deposit in our permanent digital archive, DigitalCommons@Linfield, or its successor technology. My thesis will also be available in print at Nicholson Library and can be shared via interlibrary loan.

OR

I agree to make my thesis available **only** to the Linfield College community upon its deposit in our permanent digital archive, DigitalCommons@Linfield, or its successor technology. My thesis will also be available in print at Nicholson Library and can be shared via interlibrary loan.

OR

_____ I agree to make my thesis available in print at Nicholson Library, including access for interlibrary loan.

OR

____ I agree to make my thesis available in print at Nicholson Library only.

NOTICE OF ORIGINAL WORK AND USE OF COPYRIGHT-PROTECTED MATERIALS:

If your work includes images that are not original works by you, you must include permissions from original content provider or the images will not be included in the repository. If your work includes videos, music, data sets, or other accompanying material that is not original work by you, the same copyright stipulations apply. If your work includes interviews, you must include a statement that you have the permission from the interviewees to make their interviews public. For information about obtaining permissions and sample forms, see http://copyright.columbia.edu/copyright/permissions/.

NOTICE OF APPROVAL TO USE HUMAN SUBJECTS BY THE LINFIELD COLLEGE INSTITUTIONAL RESEARCH BOARD (IRB):

If your research includes human subjects, you must include a letter of approval from the Linfield IRB. For more information, see <u>http://www.linfield.edu/irb/</u>.

NOTICE OF SUBMITTED WORK AS POTENTIALLY CONSTITUTING AN EDUCATIONAL RECORD UNDER FERPA:

Under FERPA (20 U.S.C. § 1232g), this work may constitute an educational record. By signing below, you acknowledge this fact and expressly consent to the use of this work according to the terms of this agreement.

BY SIGNING THIS FORM, I ACKNOWLEDGE THAT ALL WORK CONTAINED IN THIS PAPER IS ORIGINAL WORK BY ME OR INCLUDES APPROPRIATE CITATIONS AND/OR PERMISSIONS WHEN CITING OR INCLUDING EXCERPTS OF WORK(S) BY OTHERS.

IF APPLICABLE, I HAVE INCLUDED AN APPROVAL LETTER FROM THE IRB TO USE HUMAN SUBJECTS.

Signature Signature redacted	Date	5/	241	12	
------------------------------	------	----	-----	----	--

Printed Name Sierra noel Broussard

Approved by Faculty Advisor Signature redacted

____ Date 5/2-1/12_

Updated April 2, 2012

ACKNOWLEDGEMENTS

I would like to thank Dr. Livesay for all of her guidance. She spent countless hours advising me in this project and continually challenged me to live up to higher standards. Her excitement about psychology has been an inspiration and encouraged my love of learning.

I would also like to thank my thesis committee members Dr. Gilden, Dr. Bakner, Dr. Dunn, and Dr. Livesay for dedicating their time and support.

Finally, I would like to thank my family and friends.

Your encouragement and support has been invaluable.

TIDLE OF CONTENTS	TABL	E OF	CONT	TENTS
-------------------	------	------	------	-------

Abstract	1
General Concepts of Speech Perception	2
Confusable Phonemes	4
Motor Theory of Speech Perception	4
Amodal Theory	7
Fuzzy Logic Model of Perception	8
General Concepts of Social Priming	10
Social Priming and Automatic Processes	12
Speech Perception and Social Priming	13
Methods	14
Results	17
Discussion	17
References	22
Tables	25
Figures	27
Appendices	29

Abstract

Speech perception relies on auditory, visual, and motor cues and has been historically difficult to model, partially due to this multimodality. One of the current models is the Fuzzy Logic Model of Perception (FLMP), which suggests that if one of these types of speech mode is altered, the perception of that speech signal should be altered in a quantifiable and predictable way. The current study uses social priming to activate the schema of blindness in order to reduce reliance of visual cues of syllables with a visually identical pair. According to the FLMP, by lowering reliance on visual cues, visual confusion should also be reduced, allowing the visually confusable syllables to be identified more quickly. Although no main effect of priming was discovered, some individual syllables showed the expected facilitation while others showed inhibition. These results suggest that there is an effect of social priming on speech perception, despite the opposing reactions between syllables. Further research should use a similar kind of social priming to determine which syllables have more acoustically salient features and which have more visually salient features.

The Influence of Social Priming on Speech Perception

General Concepts of Speech Perception

Speech perception has been historically difficult to model. This difficulty arises for a number of reasons. The biggest obstacle to speech is the lack of invariance; that is while speech sources change due to differences in source, people are still able to perceive the sound. Additionally, the speech perception system doesn't rely solely on the acoustic code for speech perception; it also uses visual and motor cues to help the process.

Basic speech sounds, called phonemes, can be described based on three cues: voice onset time (VOT), place of articulation, and manner of articulation. Place of articulation identifies physically the particular area in the mouth or throat where the sound is formed. Manner of articulation describes the way the air is released; such as through the nose. Voice onset time is defined to be the length of time between when the time a consonant is released and voicing begins. Several consonants have identical placement, so they are identified by VOT or method of articulation (Nygaard & Pisoni, 1995). Despite decades of research dedicated to speech perception, researchers are still unclear which particular acoustic cues listeners are most sensitive to in the perception of any particular speech sound.

Because speech is multimodal, visual cues carry important information for the perception process. For example, information about placement of articulation is gathered primarily through the visual system, while manner and VOT are both auditory features. Risberg and Lubcker (1978) demonstrated the multimodality of speech perception by giving participants a list of words with low-pass auditory information (audio with the higher pitches filtered out), visual information, or both. Participants accurately identified only 1% of the words through visual information and correctly identified 6% of the words from auditory information. However, by

combining these inputs, subjects could identify 45% of the words. These results suggest that visual information can help ambiguous or weak auditory information. Knowing that visual information can influence auditory perception led researchers to rethink earlier views concerning speech perception.

As stated previously, a common problem for speech perception is the lack of invariance in speech, that is, while speech sounds change from person to person, our perception of them does not. The three types of variation are context, speech conditions, and speaker identity differences (Lisker & Abramson, 1967). Context-induced variation is when the phonetic environment of a sound affects its acoustic properties. For example, both typical VOT and placement of articulation can change depending on the surrounding phonemes.

Speech condition variation often refers to changes in speech rate or articulation. Because many speech sounds (particularly vowels) are mainly differentiated through temporal contrasts (length of voiced time), when speech rate changes, the same sounds become acoustically very different. However, humans are still able to perceive them as the same sounds. Articulation differences can cause some sounds to become merged together or even completely removed (Lisker & Abramson, 1967). For example, "supposed" is often pronounced as "spost" due to differences in articulation. Poor articulation is common in everyday speech; however, unless it is on the extreme end of the spectrum, these sounds are easily perceived and understood by most people.

Top-down processing also plays a key role in speech perception. Higher order processing helps fill in indistinguishable or missing sounds. Warren (1970) used white noise to cover a phoneme in a sentence and participants were able to easily understand the meaning of the sentence but were not able to identify the displaced phoneme. This finding is known as the 3

phonemic restoration effect, and highlights the top-down processing influence that context has.

Another use of top-down processing in speech perception is the stimulation of motor cues. Unless the listener is actually touching the face of the speaker, motor cues must be activated in a different way. Prior knowledge about how the mouth and throat muscles move to form speech is used for motor cue activation. Motor cues are memory based and become activated by both visual and auditory cues (Liberman et al., 1967).

Confusable Phonemes

Confusable phonemes are units of sound that are difficult to distinguish from each other (Roy & Pentland, 1998). For example, /m/ and /n/ are considered to be acoustically confusable. When phonemes have the same placement of articulation, they are said to be the same viseme, meaning they look visually similar (Lucey, Martin, & Sridharan, 2004). Pairs of voiced and non-voiced phonemes that use the same placement of articulation are always visually confusable, /ba/ and /pa/ for example. However, visemes usually encompass more than just one pair of sounds; often, there are six or more phonemes in one viseme class, meaning all of these phonemes are visually identical. For example, /t/, /d/, /z/, and /s/ are all in the /t/ viseme class. According to Roy and Pentland (1998), when people perceive speech, they process both visemes and phonemes and the second input limits the confusion caused by the other input.

Motor Theory of Speech Perception

Several important aspects of speech processing are addressed in the motor theory of speech. The motor theory of speech (Lieberman et al., 1967) suggests that people use implicit articulatory knowledge in their perception of language; meaning that when humans decode speech, they unconsciously consider the facial muscle movements necessary to make each particular sound. Lieberman et al. argued that the inclusion of visual input information to speech

4

=

perception made speech special, that is, different from the other forms of sound perception. He believed that identifying speech sounds required a different form of processing than identifying non-speech sounds, such as a bouncing ball. Categorical perception is often used as evidence of this since the visual cues of speech are a main component of the categorical divisions. However, categorical perception does not only apply to visual cues. Lisker and Abramson (1970) supported the concept by using an artificial continuum between /b/ and /p/, sounds that have the same manner and place of articulation but differ in VOT. Each step in the continuum increased VOT by 20 ms, but listeners identified the first half of the sounds as /b/ and the second half as /p/ with a distinct boundary between the categories. Additional Support for the motor theory of speech perception comes from evidence showing that the motor cortex is active when speech is processed (Watkins, Strafella, & Paus, 2003; Gow & Segawa, 2007).

Further support for the motor theory comes from the McGurk Effect (Gow & Segawa, 2009; Nasir & Ostry, 2009). In 1976, McGurk and MacDonald synced the visual cue for /ga/ with the audio cue of /ba/ and had participants report what they had heard. With only the visual cue, they perceived /ga/ and with only auditory information they perceived /ba/. But when participants received both of these cues simultaneously, they heard /da/. According to the motor theory, implicit motor cues affected the way an individual perceived the sound.

The McGurk effect is robust and occurs even when the participants are made aware of it (Galantucci, Fowler, & Turvey, 2006). Fowler and Dekle (1991) contributed even more evidence of the motor theory. Their experiment replicated the McGurk effect with conflicting auditory and motor movement; however instead of seeing a video, participants were asked to mouth a syllable while they listened to the auditory stimulus. For example, while listening to the syllable /ba/ they made the mouth movements for /ga/. Results showed that the McGurk effect was still apparent,

5

participants still perceived /da/, despite the lack of visual stimuli. These results support the motor theory because the perceptual system is depending on auditory and motor cues to perceive speech, so when they are contradictory, as in this experiment, it results in a blend of the two sounds.

There has been controversy over the validity of the "speech is special" concept. There is evidence (Galantucci, Fowler, Turvey, 2006) that indicates that speech is not special and that the motor theory applies to non-speech sounds as well. Brancazio, Best, and Fowler (2006) demonstrated this by using click-syllables that are not perceived by Americans as speech sounds. The effect still held for these even though it was not as robust as when it applies to speech sounds. Due to this evidence, the "speech is special" aspect is often not considered a part of the motor theory. Another problem for the motor theory is that the motor theory also lacks an explanation for the integration of top-down processing. For example, according to the motor theory, the phoneme restoration effect would be impossible. Because the masked phoneme would be considered a non-speech sound, it would use a different type of processing which would mean it would not include motor cues

The McGurk effect does not always result in a blend of the two inputs. Studies have shown (Jiang & Bernstein, 2011; Colin et al., 2002; Sams & Rusanen, 1998) that there are four perceptual responses: auditory correct, visual correct, fusion, and combination. Auditory correct (AC) and visual correct (VC) are when the perceptual response is completely correct with regard to auditory or visual cues respectively. Fusion is the response that people commonly associate with the McGurk effect; the two sounds are "fused" into a different sound in the way /ba/ and /ga/ become /da/. Fusions appeared more frequently when the sounds were acoustically compatible, such as when they are all voiced and have similar placement of articulation

6

(Bernstein & Jiang, 2009). Combination, as the term suggests, is the combining of two stimuli without changing either one. This reaction commonly occurs with syllable combinations that already exist in the English language, or syllables that are not acoustically similar. For example, /s/ and /l/ combine to become /sl/.

The existence of combination responses implies that the different cues remain separate for at least part of the perception process and that the perceptual process attempts to use previous linguistic knowledge to determine what sound is being perceived (Bernstein & Jiang, 2009). Combinations show that both auditory and visual components are perceived and are integrated later in processing. Additionally, AC and VC responses suggest that perception involves independent processing of auditory and visual cues.

The Amodal Theory

Rosenblum (2008) reasons that the auditory perception system is built specifically for the integration of different modes of information. He believes that thinking of perception as primarily auditory perception before non-auditory information is taken into account is a naïve way to view this complex system. He advocates considering speech perception to be an amodal function, meaning that if the perception system ignores the source of different types of information, researchers should not consider them to be different either. This allows the input of auditory, visual, and motor cues to be inseparable from the beginning of sound perception. This does not imply that the cues are not separate from each other; it implies that their original source does not matter. All information is taken in as a whole unit and perceived as a whole unit.

Similar to the motor theory, evidence to support the amodal theory comes from the McGurk effect. Studies (Rosenblum, 2005) have shown that several modifications to the McGurk stimuli are not able to remove the McGurk effect. These changes include making the

visual and auditory stimuli noticeably separate both temporally and spatially, and using auditory and visual stimuli taken from people of different genders. Rosenblum suggests that if all sources of speech were perceived independently, that these stark changes to the stimuli presentations should fail to make the McGurk effect work.

However, the amodal theory is lacking an explanation for AC, VC, and combination responses. If perceptual processing were integrated from the beginning, fusions would be the only existing responses to differing inputs. AC and VC responses are strong evidence that integration happens later in processing. The existence of combination responses guarantee that, even if AC and VC responses were due to a lack of attention to one of the cues, later integration is more likely than immediate integration.

Fuzzy Logic Model of Perception

The Fuzzy Logic Model of Perception (FLMP) depends on the fact that there are multiple sources of information being processed by the listener when perceiving a sound. The listener assimilates the information from all of the sources in order to recognize a particular sound. The sources of information consist of the many different visual, motor, and auditory cues received during speech perception (Massaro, 1987).

In the FLMP model, perceptual processes use a general algorithm and pattern recognition when recognizing words and phonemes. According to Massaro (1987) sound features go through the three stages of recognition: evaluation, integration, and decision. This model requires the brain to have a store of "prototypes," prior descriptions of each particular sound being identified. It will contain a list of features that can be compared with the currently perceived sound to determine the "goodness-of-fit." All of the prototypes should represent the paradigm of its category and have corresponding ideal values for comparison. Prototypes are generated

depending on the required task, so in general, speech perception considers individual phonemes or syllables as these prototypes.

Due to the large numbers and variety of features, it is necessary to have a standard method of measurement to determine the degree to which each matches the prototype. To achieve this, fuzzy truth values (Zadeh, 1965) are assigned to each feature. These truth values are all between 0 and 1, 0 meaning no match and 1 meaning perfect match. So a .5 rating represents a completely ambiguous feature, whereas a .8 rating represents a feature that is considerably similar to the prototype. One of the greatest advantages of the FLMP is that these ratings allow each feature to be considered on a continuum and they permit the information to be described quantifiably.

The first step of the recognition process, evaluation, determines the degree to which each sound feature corresponds to the prototype. The features are all evaluated separately and simultaneously, and then combined in the second step: integration. Feature integration creates a weighted list of all degrees of matching determined in the first stage and puts the greatest weight on the features with the least ambiguity. This way, features with the least ambiguity have the greatest influence on the final decision and features that are vague or could apply to several different sounds are basically ignored (Massaro, 1987).

In the final step of recognition, the speech perception systems use the information assembled from the previous two steps to determine which prototype fits most closely with the new sound. Each prototype is compared with the new information and evaluated for the degree to which each one matches the feature compilation created in step two. One of the most important characteristics of the model is that one feature of the sound will have the greatest impact on decision when the other features are the most ambiguous.

9

The FLMP is ideal for explaining speech perception because, unlike the motor theory, it allows for both bottom-up and top-down information to be included in the evaluation stage, which are both vital components of the speech perception process. Experiments have investigated both kinds of information with regards to this model (Massaro, 1987). Information has included auditory and visual factors as well as syntactic and semantic features. The FLMP also allows for quantifiable predictions. Even though we aren't able to know exactly what the fuzzy truth values for each sound are, they can be discussed hypothetically.

General Concepts of Social Priming

Behavior can be influenced when factors in the environment encourage a certain schema (Bargh, Chen, & Burrows 1996). Schemas are preconceived mental themes that help people organize social structures. Unlike semantic priming, where very specific shared semantic features relate stimuli, social priming triggers entire schemas that can affect behavior on a much larger scale. For instance, instead of just causing the participant to equate 'doctor' and 'nurse' because of shared semantic features, social priming can activate the participant's schema for doctor's offices to alter their behavior to match the way they would behave in that situation. Because priming is implicit, people are unaware of their own behavior change. It is often necessary for the participant to be unaware that they are being primed in order for the priming to affect their behavior. However, Neely (1977) showed that reactions less than 500 ms are automatic, so in that brief period of time, people cannot influence their behavior even if they are aware of the priming.

Bargh, Chen, and Burrows (1996), primed participants with an old age schema by asking participants to unscramble sets of words to create grammatically correct sentences. Many of these sentences had words relating to old age stereotypes like "bingo," "slow," and "Florida."

10

Thus participants would be implicitly exposed to the old age stereotypes so that their corresponding schema was activated. After the participants left the lab, a confederate secretly timed how long it took them to walk down the hallway to the exit. Results showed a significant decrease in average walking speed by participants who had received the age primes. Bargh et al. (1996) suggested that this was because slowness is an aspect that most people share in their schema of old age.

Bargh et al. (1996) showed further support for social priming in another experiment using words related to politeness such as "respect," "honor," and "yield," or words relating to rudeness such as "aggressive," "rude," and "disturb." After completing the scrambled sentence task, participants were lead out of the room and told that the next task would be explained momentarily. While the participant waited, the experimenter became engaged in a conversation with a confederate that continued until ten minutes passed or the participant interrupted them and asks to be given the next task. In the politeness condition 18% of participants interrupted, whereas 65% interrupted in the rudeness condition. The neutral condition yielded 37% participant interruptions. There was also a difference in the amount of time the participants took to interrupt the experimenter, with the participants in the politeness condition taking more time and the participants in the rudeness condition taking less time. So by activating either a rude or polite schema, participants' behavior changed drastically.

Other studies have also demonstrated the robustness of the social priming effect. Garcia et al. (2002) investigated the effect of social priming on the bystander effect. The bystander effect is the decrease in likelihood of an individual's helping behavior as the number of bystanders increase, and studies have shown this effect to be strong and that most people do not predict they will be affected by it (Darley & Latané, 1968). Garcia et al. (2002) gave participants

11

a questionnaire designed to make them to imagine being at dinner with friends and then were asked how much of their income they would donate to charity. Participants who imagined many friends said they would give less money than participants who imagined few or no friends. These results correspond with the way participants behave when there are actually people present. So activating a crowd schema causes participants to react in the same way they do in an actual crowd.

In the previous experiments, the expected result was very clearly integrated with the basic schemas for politeness and old age. Garcia et al. (2002) showed that social priming does not only affect behaviors that are obviously part of a specific schema, but can affect behavior through parts of schemas that we are not explicitly aware are related.

Social priming and automatic processes

Studies have shown that when one person imitates another, the person being imitated has increased feelings of closeness, altruistic behavior, and trust. It also aids in the smoothness of a social interaction. Leighton et al. (2008), they used social priming to determine whether imitation and pleasant behavior is bidirectional. They hypothesized that if participants are primed with pro-social words such as "friend," "join," and "ally" they will be more inclined to use imitation in their next social encounter. Similarly, if primed with anti-social words such as "single," "apart," or "distrust", they will be less likely to imitate their partner in their next social interaction. They were then directed to imitate hand gestures, either open palm or closed fist, which would appear on the screen. Participants who received the anti-social prime were significantly faster. This study demonstrates that social priming can alter reaction times, something which people have limited control over.

Goldfarb, Aisenburg, and Henik (2010) used the Stroop (1935) task (a task that requires participants to look at a color word such as "red" written in a different colored ink, and say the color of the ink) to investigate the effects of social priming on automatic processes. In the Stroop task, participants have to suppress an automatic process, reading, in order to perform a color naming task. The suppression causes interference, which results in longer reaction times for the color naming. To briefly inhibit participants' automatic reading abilities, Goldfarb et al. primed the social construct of dyslexia. Participants who were primed with dyslexia were significantly faster at the Stroop task, compared to the non-primed individuals. This suggests that social priming can influence automatic processes.

Speech Perception and Social Priming

The current study aims to reduce the confusability of phonemes in the same viseme (visual cue) category through social priming of the concept of blindness. All of the models agree that speech perception relies on auditory and visual information. Therefore, if the perception of visual cues is changed, it should alter speech perception as a whole. Since speech perception is an automatic process and social priming has been shown to affect automatic processes, it should be possible to alter speech perception using social priming. As shown the experiment by Goldfarb et al. (2011), automatic processes are extremely difficult to change through explicit control. Activating a schema for blindness is ideal for altering the reliance on visual cues because this schema has a clear relationship with decreased use of the visual system.

Decreasing the reliance on visual processing part of the speech perception system should cause visual cues to be more ambiguous. According to the FLMP (Massaro, 1987), these cues will then be given less weight in the integration step. So if visual information had previously been the main defining feature of a particular speech sound, this information will now be ignored

13

in favor of the second most defining feature. It should also make some sounds more easily identified. For example, if there are two types of heavily weighted, conflicting information the visual component will be less heavily weighted, leaving only one kind of information. If we assume that this decreases decision speed, and thus processing speed, this should be shown through shorter identification reaction time.

If participants are primed with the concept of blindness and then presented with a mix of confusable and non-confusable syllables, they should have a shorter RT when identifying confusable syllables than if they had not received the priming. Furthermore, the blindness priming should not have an effect on the non-confusable syllables and the non-confusable syllables should have a shorter identification RT than confusable syllables in general.

Methods

Participants

Thirty-eight native English speakers from a small liberal arts college aged 18 to 22 participated in this study. Most were enrolled in introductory psychology classes and received class credit and entrance into a drawing with a cash prize for their participation. All participants reported corrected or corrected-to normal vision and normal hearing. Each was assigned to one of four lists.

Materials

The blind and low-vision prime words were created using an online Latent Semantic Analysis Near Neighbors generator provided by CU Boulder. The final list of 20 words (See Appendix B) was compiled from the top selections gathered by running the analysis on the word "blind" and the phrase "vision impairment." Sentences were generated manually by selecting four other words that were not found on the lists. The 20 non-prime sentences were created in a similar manner, but words from the prime list were not used for any of the five words in each sentence.

All of the video clips were filmed with a Kodak Playsport Zx5 HD Pocket Video Camera. Videos were of the lower portion of an 18 year old male volunteer's face. The volunteer was a theatre arts major and was trained in diction. Each clip was trimmed at the beginning and end to make the timing of each clip as similar as possible and to remove excessive video around the desired stimuli. All video clips are immediately followed by a 100 millisecond tone generated by a sound file editing program called Audacity. Syllables were chosen based on a table of visually confusable and non-confusable phonemes (Lucey, Martin, & Sridharan, 2004). Five pairs of visually confusable syllables were selected, each pair consisting of a voiced an unvoiced phoneme: /ba/, /pa/, /ta/, /da/, /fa/, /va/, /sa/, /za/, /ka/, and /ga/. Three non-confusable syllables were also chosen: /ha/, /wa/, and /ya/. Each syllable had its own video clip and was repeated three times.

Four lists were created from the two types of stimuli. Because there are two independent variables (syllable confusability and presence of blindness words), four lists were created for counterbalancing purposes. Each list is made up of a block of sentences, all the video clips, another block of sentences, and all of the clips again. The first block of sentences is always the opposite of the second block. For example, List 1 consists of neutral sentences first, followed by prime sentences. The orders of the syllable videos were randomized in each list.

The experiment was run using SuperLab and all auditory reaction time responses were recorded by a Cedrus SV-1 Voice Key. Gain and delay were adjusted according to uncontrollable external conditions and varied slightly for each participant. Participants were allowed to adjust volume to their comfort, but only three chose to do so and these adjustments 15

were negligible.

Procedure

Participants came to the lab and completed an informed consent form and were given a randomly designated list. Then they were directed to sit at a computer and put on the Voice Key headset, adjusting the microphone for optimal performance. After this, a brief demographic questionnaire was presented (See Appendix A). Once participants completed the demographic questionnaire, they were given a set of practice trials for the video task. Each trial was composed of a single syllable, repeated three times. None of the syllables in the practice trials are used as stimuli in the testing part of the experiment.

Following the practice trials, participants were given instructions and an example for the sentence task. They were presented with a line of five words in random order. They were told to choose four out of the five available words, create a grammatically correct sentence with them, and type each sentence into the computer using the keyboard. After completing 20 sentence trials, participants began the video task. The instructions directed participants to watch the video clip and, after hearing a brief tone, repeat aloud the syllable they heard as quickly as possible. Reaction times were recorded for all responses, and if the response was longer than 2000 ms, it was considered a no response and reaction times of less than 200 ms were considered technical errors by the computer. As soon as the computer detected the participants' voice, the program instructed them to type the two letter syllable they heard. After this was performed for all 13 syllables, another sentence task was presented. This task was identical to the previous one but contained different stimuli (i.e. if the previous sentences were neutral, these sentences would contain the words relating to blindness). A second video task, identical to the original, followed after the participant completed all of the sentence trials. The participants were then informed that

the experiment was over and orally debriefed by the experimenter.

Results

Analyses focused on participant's reaction time (RT) to the responses to the syllable videos. Data for error and non-response trials were excluded (less than 3% of all responses) and two participants were left out of analysis due to over half of their trials consisting of errors or non-responses. After this, data was cleaned for response times that were ± 2.5 SD from each participant's mean, affecting less than 2% of total responses. Please see table 1.

An initial 2 (confusability of syllable) x 2 (presence of prime) repeated measures ANOVA was calculated. A marginal main effect of syllable confusability was found, F(1, 34) =3.641, p < .068. No other main effect was found, F(1, 34) = 1.202, p < .154; nor was there a significant interaction F(1, 34) = 1.138, p < .260.

A by syllable analysis was performed to investigate the influence of blindness priming or individual syllables. Please see Table 2 for means and SD. If a trial was missing its pair, that trial was ignored, thus not all degrees of freedom are the same. Paired sample t-tests were performed for each syllable. Two syllables, /ba/ and /ga/, show significant facilitation, $t_B(29) = 2.29$, $t_G(30)$ = 2.328, meaning RT was shorter after the priming sentences compared to RT after the neutral sentences.

Two syllables, /fa/ and /wa/, showed interference, $t_F(29) = -3.016$, $t_W(26) = -2.787$, implying that RT was longer after the neutral sentences. No other syllable comparisons were significant.

Discussion

The original hypothesis was not supported; there was no main effect of priming overall and no interaction. The lack of a main effect is possibly due to the amount of variation in the

_

reaction time responses. Increasing the number of participants, thus increasing power, might show a significant main effect of priming.

However, a closer inspection of the data revealed some interesting by syllable effects. The syllables /ba/ and /ga/ were facilitated when preceded by the blindness primes, whereas /fa/ and /wa/ showed interference. Why would the blindness primes affect each syllable differently? There are several possible answers, but it is important to first note that /wa/ might be a special case. First, /wa/ was a non-confusable syllable. More importantly, a significant number of /wa/ trials were considered "no response" because the participant took more than 2000 ms to respond. I believe this was due to a technical difficulty with the microphone failing to pick up the speech sound because of way the sound is formed. The lack of total trials might have caused /wa/ to have a greater variability, which was already high overall, so the inhibition of /wa/ can be discounted as a technical error.

The facilitation of /ba/ and /ga/ was the predicted effect for all syllables, but /fa/ is being affected in the opposite direction. This difference could be explained in two ways. One of the obvious differences between the facilitated syllables and /fa/ is that /b/ and /g/ are voiced phonemes while /f/ is not. This could mean that people are more dependent on visual information for /f/ than they are for /b/ and /g/ because the voiced syllables have more distinctive auditory information. Conversely, different effects of priming could be due to the distinctiveness of the visual cues. The phonemes /b/ and /g/ each share their viseme with many other phonemes (high viseme neighborhood density) while /f/ is only visually identical to /v/ (low viseme neighborhood density). Low viseme density should lead to the visual information of /f/ should be much more salient than the visual information for /b/ and /g/. If this is true, then according the FLMP (Massaro, 1987), the main feature of /f/ would be the main factor contributing to the

18

identification of the sound. However, when the sound is preceded by blindness prime words, participants should be less reliant on their visual cues. This means that /f/ becomes more ambiguous because it is being identified mainly through its less distinctive auditory cues. On the other hand, /b/ and /g/ are mainly distinguished through auditory cues, so making participants less reliant on visual cues only serves to make the visual cues more ambiguous and easier to ignore using the FLMP, thus facilitation.

It is unclear why only these three syllables were significantly affected by the blindness priming words. One explanation is that the other syllables rely on visual and auditory cues more equally, so disregarding the visual aspects of the sounds doesn't have much effect on identification because the auditory cues are distinctive enough to allow quick decisions without being so distinctive that visual information is unhelpful. Also, low power may have caused some of the other voiced consonants to fail to show significant facilitation.

Although the results did not support the hypothesis, the differing effects of priming between syllables give stronger evidence for the FLMP than the original hypotheses. According to the amodal theory, social priming should be unable to effect speech perception. If the source of the cue doesn't matter, lowering dependency on one of the sources should have little effect on speech processing over all. However, the priming could take affect after the cue has been initially processed. In this case, there should be no effect because the information cannot be disassociated at this point.

On the other hand, the motor theory does support the idea that socially priming blindness could alter speech perception; however, the motor theory can't make quantifiable predictions about how social priming would affect it. Thus, the motor theory can't explain why /fa/ was inhibited. It would suggest that the identification of any syllable with some type of confusing

visual information would be facilitated by socially priming blindness. The FLMP shows that this might not be the case: if the visual information is essential in identifying a sound, removing this information will inhibit its identification. Because of these features, the FLMP appears to be the ideal model for explaining the results of this experiment.

So what does this tell us about speech perception in general? Since the results seem to support the FLMP, it suggests that speech is multimodal, and that cues from these different modes are integrated late enough in the process to be altered by schema activation. Furthermore, it suggests that altering one of these modes causes the perception of speech to be altered in a quantifiable and predictable way. Each syllable reacts slightly differently to each alteration due to whether it has more salient visual cues or more salient acoustic cues. So according to the results, speech perception is a type of automatic decision making process.

It is also important to consider what information these results give us about social priming. Since the experiment showed that priming can influence certain syllables and we are constantly activating schemas in daily life, what kinds of influences are these schemas having on or perception of speech? One answer is that we likely use schema to help nullify some of the variance in speech, particularly variances such as accent. For example, if someone has a schema for Texas, part of that schema will tell him or her to expect different pronunciations of words, especially vowels. When they go to Texas, the schema is activated, temporarily changing their prototype for certain sounds, making it faster and easier to identify them.

An obvious issue to discuss is that confusable syllables cannot be classified a simply "confusable." The basic premise of the experiment assumed that any syllable with a visually identical mate is confusable. The results suggested that this is not true, so putting them all into one group of "confusable syllables" is not an effective way of defining them because the

confusion does not all come from the same source. It would be more helpful to determine which sounds are visually confusable but acoustically distinct and which sounds are acoustically confusable but visually distinct.

Future research should explore the differences in how blindness priming affects these two types of confusable syllables. The phonemes /m/ and /n/ would be ideal to use as acoustically confusable but visually distinct sounds. Comparing the effect of priming on these syllables with the effect on /b/ and /g/ would hopefully increase the understanding of which cues people rely on for each piece of speech sound as well as move toward being able to predict cue reliance in speech as a whole.

=

References

- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71 (2), 230-244.
- Bernstein, L.E., & Jiang, J. (2009). Visual speech perception, optical phonetics, and synthetic speech. In A. Liew & S. Wang (Eds.), Visual Speech Recognition: Lip Segmentation and Mapping. IGI Global, 439-461.
- Colin C., Radeau M., Soquet A., Demolin D., Colin F., & Deltenre P., (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology*, 113, 495–506.
- Darley, J. M. & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383.
- Fowler, C., & Dekle, D. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. Journal Experimental Psychology: Human Perception and Performance, 17, 816-828.
- Galantucci, B., Fowler, C., & Turvey M. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13, 261-377.
- Garcia, S. M., Weaver, K. D., Moskowitz, G. B., & Darley, J. M. (2002). Crowded minds: The implicit bystander effect. *Journal of Personality and Social Psychology*, *83*, 843–853.
- Goldfarb, L., Aisenburg, D., & Henik, A. (2010). Think the thought, walk the walk Social priming reduces the Stroop effect. *Cognition*, 118, 193–200.
- Gow, D. & Segawa, J. (2009). Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *Cognition*, 110, 222-236.

- Leighton, J., Bird, G., Orsini, C., & Heyes, C. (2010). Social attitudes modulate automatic imitation. *Journal of Experimental Social Psychology*, *46*(6), 905–910.
- Lieberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* 74, 431–461.
- Lisker, L. & Abramson, A. (1967). Some effects of context on voice onset time in English plosives. *Language and Speech*, *10*(1), 1–28.
- Lucey, P, Martin, T, & Sridharan, S. (2004). Confusability of phonemes grouped according to their visemes in noisy environments. 10th Australian International Conference on Speech Science & Technology. 265-270.
- Massaro, D. W. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Lawrence Erlbaum Assoc, Inc.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 106, 226-254.
- Nygaard, L & Pisoni, D. (1995). Speech perception: New directions in research and theory. In J.
 Miller, and P., Eimas. *Handbook of Perception and Cognition: Speech, Language, and Communication*. San Diego: Academic Press.
- Rosenblum, L.D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405-409.
- Roy, D & Pentland, A. (2002). Learning words from sights and sounds: A computational model. Cognitive Science, 26(1), 113-146

Sams, M. & Rusanen, S. (1998). Integration of dichotically and visually presented speech

stimuli. Audtitory-Visual Speech Processing.

Stroop, J. R. (1935). Studies in interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–661.

Warren, R. (1970). Restoration of missing speech sounds. Science 167 (3917): 392-393.

Watkins, K., Strafella, A., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 48, 989–994.

Zadeh, L. (1965). Fuzzy sets. Information and Control, 8, 338-353.

Ξ

Running head: SOCIAL PRIMING AND SPEECH PERCEPTION

Tables

Table 1

Comparison of Mean Reaction Time for Confusability and Prime Type

Nei	ıtral	Pri	ime	
М	SD	М	SD	Total Mean*
825.51	19.89	829.78	15.40	827.49
786.90	25.96	816.70	16.93	801.80
806.21		823.09		<u> </u>
	Neu M 825.51 786.90 806.21	Neutral M SD 825.51 19.89 786.90 25.96 806.21 1000000000000000000000000000000000000	Neutral Print M SD M 825.51 19.89 829.78 786.90 25.96 816.70 806.21 823.09	Neutral Prime M SD M SD 825.51 19.89 829.78 15.40 786.90 25.96 816.70 16.93 806.21 823.09 1000 1000

*marginally significant, p = .64

=

Table 2

Descriptive Statistics for Individual Syllable Reaction Times by Prime Type

(Significant Differences Only)

Syllable	M_N	SD_N	M_P	SD_P
Ba (n=30)	846.30	225.72	798.43	156.69
Ga (n= 31)	811.61	186.03	890.68	225.63
Fa (n=29)	821.97	139.20	915.45	158.88
Wa (n=27)	757.00	151.73	851.00	185.95













Ξ.

Running head: SOCIAL PRIMING AND SPEECH PERCEPTION

Appendices

Appendix A

What is your gender? М F How many years have you been in college? (Rounded up to the nearest whole number.) How old are you? Do you have noticeable hearing loss? YES NO Do you wear glasses or contacts? YES NO Do you have normal or corrected-to-normal vision? YES NO Are you wearing your glasses or contacts? YES NO N/A

=

Defect	Disease		
Eye	Disability		
Shades	Impair		
Darker	Deaf		
Headache	Handicapped		
Sensitive	Corneas		
Opaque	Damage		
Vision	Syndrome		
Blind	Exposure		
Abnormal	Afflicts		

Appendix B

,

=