There are times in which our speech production processes falter, and the valuable mechanisms we generally take for granted temporarily fail to function. The mechanisms underlying speech production, particularly word retrieval, have not been clearly established. The fluency of speech can be influenced by many factors, including nonverbal variables such as the one of primary interest in the current study: the ability to produce gestures while speaking (e.g., Frick-Horbury & Guttentag, 1998; Morsella & Krauss, 2004). The study of how gestures impact language production is of potential importance to theory development because existing theories of language production (e.g., MacKay, 1987) would need to explain any obtained effect of gesturing. Additionally, if gesturing is determined to be an effective way to improve speech fluency, gesturing could be incorporated into certain types of speech therapy.

Although speech generally flows quite fluently, a number of speech disfluency types exist. Fluency measures can include counts of stutters, incomplete words, and repeated words or phrases, as well as assessments of pause length or pause frequency, and the production of filler words (e.g., um, uh, er) while searching for a particular word or planning utterances (e.g., Burke & Shafto, 2008; Kasl & Mahl, 1965; Mahl, 1987), but as noted by Kemper, Herman, and Lian (2003), fluency measures have varied dramatically across studies. Most people have experienced the frustrating sensation in which they know a word they are looking for, but temporarily cannot access it. These retrieval problems are called tip-of-the-tongue (TOT) states, and they are experienced by people of all ages (Burke, MacKay, Worthley, & Wade, 1991; Maylor, 1990). TOTs occur when the phonological components of a word are not accessible even though the semantic information about the word has been retrieved (James & Burke, 2000). TOTs therefore represent a failure of lexical access that disrupts speech fluency. However, they are not the only type of speech disfluency, and there are numerous ways in which speech

**ABSTRACT.** Evidence has indicated that gestures can impact speech production. Specifically, restricting participants’ gestures can negatively impact the production of individual words and the fluency of connected speech (e.g., Frick-Horbury & Guttentag, 1998; Morsella & Krauss, 2004). We manipulated participants’ ability to gesture while describing pictures with active or static content. We predicted that prohibiting gestures would decrease speech fluency, especially for pictures with active content. We measured the amount of time participants paused during speech, the rate of speech fillers, and the occurrence of obvious tip-of-the-tongue instances during the descriptions. Neither gesture condition nor picture content yielded main effects on speech fluency, but allowing gestures reduced the production of speech fillers in descriptions of pictures with static but not active content, $F(1, 21) = 6.34, p = .02, \eta^2_p = .23$, contrary to prediction. Findings have suggested that being able to gesture may slightly impact fluency even if gestures are not actually produced while speaking.
can be disrupted, with a variety of underlying causes. For example, Mahl (1987) described several types of speech disfluency that vary with the speaker’s anxiety level, and Kemper et al. (2003) described ways in which normal aging affects the fluency of speech.

One specific branch of research on factors contributing to the fluency of speech production has tested the effect of body (particularly hand) movement on tasks involving the retrieval of specific target words. Frick-Horbury and Guttentag (1998) found that more words were correctly produced by participants whose hand movements were unrestrained than by participants in a restrained group. Interestingly, this benefit of being able to gesture occurred even though correct answers were generally produced without accompanying gestures. In other words, having the ability to gesture made word retrieval easier, although actually gesturing did not appear to impact task performance. Ravizza (2003) found that finger tapping movements with no obvious meaning improved lexical production compared to remaining still. This effect was limited to resolution of TOT states (i.e., coming up with the correct word once in a TOT state), as movement did not help generate correct responses when the speaker was not in a TOT state (see also Beattie & Coughlan, 1999).

In addition to affecting TOT states, prohibiting gestures can impact speech fluency during utterances longer than a single word. Graham and Heywood (1975) tested speech fluency when participants were either free to move or told to fold their arms while describing line drawings. Movement restriction did increase the difficulty of speech production as indicated by one of several measures of pause time, but did not affect other measures of pause time, production of hesitations (e.g., uh’s), or speech rate. Hostetler and Skirving (2011) found that participants gestured more when describing an event after watching a cartoon that included verbal narration than when only hearing the narration (with no visual component). However, they found no difference in the number of speech disfluencies produced in the two conditions, suggesting that the increased gesturing in the visual condition resulted from accessing mental imagery rather than from problems with lexical access. Nevertheless, visuospatial content descriptions did prompt more gestures than verbal content ones. Morsella and Krauss (2004) examined the use of gestures during descriptions of objects, and found that descriptions of objects removed from view were accompanied by more gestures than descriptions of objects that remained visible. There was a noticeable use of gestures even in descriptions of stimuli that were present during the task, suggesting that gestures may serve as a method of aiding lexical retrieval, in apparent contradiction to the results of Hostettler and Skirving (2011). Morsella and Krauss (2004) also found that speech rate decreased when instructions prohibited gesturing, regardless of whether or not the stimulus was present, reflecting participants’ reduced speech fluidity when restrained from gesturing.

A critical concern about these previous studies of gesture and discourse, in addition to the TOT research described previously, is that participants in the nongesturing group might have been distracted by having to focus on not moving or making gestures, leading to the obtained performance decrements. An experiment by Rauscher, Krauss, and Chen (1996) strongly influenced the design of the current study because they planned their study to avoid the confounding variable of distraction. Rauscher et al. (1996) showed cartoon video clips and compared participants’ speech fluency while describing the videos under conditions that allowed or prohibited gestures. In order to prevent participants from discovering that the focus of the research was gesturing, the researchers attached electrodes to either participants’ ankles (the gestures-allowed condition) or palms (the gestures-prohibited condition) and told participants they were collecting physiological measures through the electrodes. In reality, the electrodes were attached to prevent participant movement. Participants were told to keep their limbs in contact with the electrodes (i.e., to not move). They calculated the number of filled pauses, silent pauses, truncated sentences, word fragments, and repeated words in the video descriptions and found that being allowed to gesture specifically benefited speech fluency when participants were describing spatial content. In other words, when participants were describing the location of an item (e.g., when using prepositional phrases beginning with in or on), they produced fewer pauses and other disfluencies in the gestures-allowed condition than the gestures-prohibited condition. When participants were describing nonspatial content, there was no benefit of being able to gesture.

The Current Study
In this initial, pilot study of the effect of gesturing
on oral picture descriptions, we tested whether the ability to gesture facilitates lexical retrieval and affects the overall fluency of speech production. Critically, we eliminated the confounding variable of distraction or divided attention found in many past experiments (i.e., in which participants in the restrained condition were specifically told to remain still) by using methodology similar to Rauscher et al. (1996). We placed electrodes either on participants’ palms or ankles, and all participants were told we were measuring their skin conductance to assess their anxiety. In reality, the electrodes served to ensure that half the participants (those with electrodes on their palms) would not gesture whereas the other half (those with electrodes on their ankles) were free to gesture. This allowed us to avoid disclosing our interest in gestures and to avoid telling participants in one group to remain still and not gesture. This ensured that both groups had a similar experience except for the placement of the electrodes, thereby allowing a more accurate assessment of the effect of gesturing on speech production than provided by many previous studies.

Because Rauscher et al. (1996) and Hostetter and Skirving (2011) found differences in the effect of allowing gestures when describing spatial versus nonspatial content, we manipulated picture content to test whether active and static content yield similar gesture use and produce similar benefits of gestures on speech fluency. By using pictures as stimuli to elicit descriptions, we expected frequent use of gestures for those given the freedom to do so. Speech fluency was assessed through three different measures: speech fillers (e.g., uh, um), the percent of time during which the speaker paused while producing each description, and identifiable TOT states. We predicted that being allowed to gesture would benefit speech fluency overall, with a pronounced benefit for descriptions of pictures with active compared to static content. In other words, we expected fewer speech fillers, less time spent pausing during speech, and fewer identifiable TOTs in the gestures-allowed than in the gestures-prohibited condition, and we expected the benefit of being allowed to gesture would be greater for descriptions of pictures with active content than for descriptions of pictures with static content.

Method

Participants
Participants were 23 undergraduate students ranging in age from 18 to 46 years (M = 25.65, SD = 8.79). Participants completed between 12 and 16 years of education (M = 14.39; SD = 1.31). There were 17 women and 6 men. Eighteen participants identified themselves as White, two as Hispanic, one as Black, one as Asian or Pacific Islander, and one as American Indian or Alaskan Native. All were fluent speakers of English with a mean of 29.75 items correct out of 40 (SD = 5.00) on the Shipley Vocabulary Test (Shipley, 1940). One participant identified herself as bilingual. An additional two participants were tested but their data were not analyzed (one was not fluent in English and another misunderstood task instructions). All participants were compensated with credit to be used in their choice of psychology course.

Materials
Stimuli consisted of six pairs of pictures matched on the type of visual content, with one member of each pair having active content and one having static content. For example, one picture included players on a lacrosse field in the middle of a game, and its paired picture included a person posing for a lacrosse team picture without any action involved. Another picture included a wrecked car sitting in a parking lot, and its paired picture included a car in midair during a crash on a racetrack. Pictures were presented one at a time via a PowerPoint slideshow, with five randomly ordered versions created to minimize order effects.

Participants sat in a chair, and nonfunctional electrodes were attached to either their palms or ankles. The electrodes were connected to a nonfunctional machine, ostensibly for measuring skin conductance. A Sony Handycam® (model DCR-DVD308) video recorder and an Optimus™ (model CTR-111) audio recorder were used to record participant gestures and speech.

We administered an informed consent form, and a demographics questionnaire on which we asked for each participant’s age, sex, years of education completed, race or ethnicity, and native language(s). We also gave the 40-item, multiple-choice Shipley (1940) vocabulary test.

Procedure
Our project was approved by our university’s institutional review board. Participants read and signed the consent form and then completed a demographics form and the vocabulary test. Participants received instructions that skin conductance measurements would be taken through either their palms or ankles.
Participants were assigned to conditions in alternating order of arrival for participation, resulting in 12 participants in the gestures-prohibited condition and 11 in the gestures-allowed condition. Participants in each condition were treated identically except for the placement of the electrodes. In reality, the skin conductance machine was not functional, but this deception allowed participants to be placed in either the gestures-allowed (electrodes on ankles) or gestures-prohibited (electrodes on palms) condition without the confounding variable of distraction (i.e., both groups were potentially distracted by having electrodes attached to them). The experiment thus avoided demand characteristics that could come with participants knowing that gesturing was a component of the experiment. To be sure that gesturing was actually prevented, the limbs with the attached electrodes were also loosely tethered to the chair by easily removed restraints. Participants were told not to move their hands (in the gestures-prohibited condition) or their feet (in the gestures-allowed condition). They were led to believe that we did not want them to move in order to ensure that they did not accidentally dislodge the electrodes and prevent data collection.

Each participant described all 12 pictures. In other words, picture content (active versus static) was manipulated within participants and was not presented in blocked fashion. Participants were instructed to look at each of the 12 pictures on the laptop screen and describe only the physical content of the pictures (not their emotional responses to them) in as much detail as possible. They were told to aim for at least 1 min of speech to thoroughly describe the scene as if talking to someone who could not see the picture. Participants were to indicate to the experimenter when they were ready to proceed to the next slide. Participants were videotaped and audio-recorded for later coding of their gestures and speech. When participants had completed their descriptions, they were debriefed and told the real purpose of the electrodes. Participants were compensated with course credit following completion of the study.

To analyze participants’ speech production, one experimenter listened to and transcribed each audio recording, and then a second experimenter listened to the same audio recording and double-checked and edited the original transcripts. Then the first experimenter went back and approved or rejected any changes made by the second experimenter. The first experimenter then counted the number of words and the number of speech fillers, and these were also double-checked by the second experimenter. The first experimenter assessed gestures from each video recording, using a broad definition of gesturing (i.e., any motion of the hands and/or arms during the speech task), and the identified gestures were confirmed by a second experimenter. We divided the number of speech fillers by the number of words spoken for each description to get a percentage of fillers in each condition. We computed the percent of time speakers paused during each picture description by dividing total amount of pause time by total speaking time in each condition. Finally, we tallied instances of identifiable TOTs and the number of gestures produced in each condition.

Results

We conducted 2 x 2 mixed factorial analyses of variance (ANOVAs) with gesture condition as a between-participant variable, and picture content as a repeated measure. Table 1 (top) displays the means and standard deviations for speech fillers in each condition. There were no main effects of gesture condition, $F(1, 21) = 0.98, p = .33, \eta^2_p = .05$, or of picture content, $F(1, 21) = 0.28, p = .61, \eta^2_p = .01$, on the frequency of speech fillers. However, gesture condition interacted with picture content, $F(1, 21) = 6.34, p = .02, \eta^2_p = .23$. Participants in the gestures-allowed condition produced a similar percentage of speech fillers for active and static pictures, $t(10) = 1.23, p = .25$. However, participants in the gestures-prohibited condition produced more speech fillers when describing static than active pictures, $t(11) = 2.51, p = .03$.

Table 1 (bottom) displays the means and standard deviations for percent of time spent pausing

<table>
<thead>
<tr>
<th></th>
<th>Speech Fillers (as % of Words Produced)</th>
<th>Pause Time (as % of Time Spent Speaking)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gestures Allowed</td>
<td>Gestures Prohibited</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Speech Fillers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Content</td>
<td>4.4%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Static Content</td>
<td>3.9%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Percent of Time with Pauses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Content</td>
<td>29.9%</td>
<td>13.6%</td>
</tr>
<tr>
<td>Static Content</td>
<td>31.6%</td>
<td>14.6%</td>
</tr>
</tbody>
</table>
in each condition. There was no main effect of gesture condition, $F(1, 21) = 0.05$, $p = .83$, $\eta^2_p < .01$, or picture content, $F(1, 21) = 2.55$, $p = .13$, $\eta^2_p = .11$, and no interaction between gesture condition and picture content, $F(1, 21) = 0.21$, $p = .65$, $\eta^2_p = .01$.

Of 23 participants in the study, only two people appeared to experience one TOT state each, as indicated by expressions of frustration and excited resolution statements indicating that they knew the word but had temporary difficulty retrieving it. One example is the statement, “... maybe a... hahahaha what’s it called? Um... it was black. An El Camino! That’s what it looks like, an El Camino.” The participant came up with the right name for the car, but it took the speaker effort to retrieve the word. Both of the observed TOT states occurred in the gestures-prohibited condition.

In addition to finding few clear examples of TOTs, we also found that very few gestures were produced during the description task. Of 11 participants in the gestures-allowed condition, only two made any gestures. One participant performed one distinct gesture (while describing a cheerleader doing the splits in the air, the speaker put her hands together with fingers pointed out in opposite directions to create a ‘v’ to demonstrate how the cheerleaders’ legs were positioned in the splits), and the other produced one distinct gesture and one questionable hand motion that could be interpreted as a gesture. Thus, even when participants experienced no restriction in ability to move their hands, few took the opportunity to gesture.

**Discussion**

We used three different measures of speech fluency in this pilot study on the impact of gestures during a picture description task. On the whole, being allowed to gesture did not dramatically benefit speech fluency. Gesture condition interacted with type of picture content being described for our measure of speech fillers, but not for the amount of time spent pausing. There were also too few clear instances of TOT states to make meaningful conclusions regarding this measure, although the finding that both clear TOTs occurred for participants in the gestures-prohibited condition suggests that limiting gestures could increase TOT states.

The finding that the ability to gesture only reduced speech fillers in the static condition was contrary to our prediction. When participants described pictures, such as an airborne car crashing or people diving for the ball in lacrosse, the freedom to gesture did not lead to the production of more fluent speech. However, when describing a picture of a stationary car, or a girl in a lacrosse uniform standing still, the ability to gesture did lead to the production of more fluent speech. One possible explanation for this unexpected finding is that many of our pictures with static content had fewer detailed components for participants to describe than our pictures with active content. For example, in one static picture there was a solitary man in a karate uniform standing against a dark background, and the paired active picture contained two men engaged in martial arts competition with a house and other details visible in the background. Although both pictures did contain many details, most of our participants described the pictures without including every possible specific detail. This unintentional difference in visual complexity between picture types might have left people to utter *um* and *uh* while searching for something more to say about the static pictures, but only in the condition in which gestures were prohibited. Finding a main effect of picture content on speech filler production would indicate a greater number of speech fillers produced for active versus static pictures. If this were the case, differences in visual complexity between the two conditions could have affected the outcome. However, we did not obtain a main effect of picture content on speech filler production, so the possible confounding of visual complexity with picture content is not a substantial concern. Further, it does not reduce the importance of the obtained interaction between gesture condition and picture content: The reduced speech fillers in the static condition for participants who were allowed to gesture indicates that they had less difficulty finding something to describe about the pictures than participants who were not allowed to gesture, an interesting finding consistent with the idea that gestures can facilitate language production. Nevertheless, future research should aim to control the number of visual elements contained within pictures in the two conditions.

Our lack of a main effect of gesture condition on the production of filler terms or proportion of speech time spent pausing is in accordance with some (e.g., Beattie & Coughlan, 1999; Hostetter & Skirving, 2011; Ravizza, 2003) but not all past research (e.g., Frick-Horbury & Gutten tag, 1998; Pine, Bird, & Kirk, 2007). Our study eliminated the potential distraction of participants in one condition being told to remain still, suggesting that previous findings might have been due to the divided attention necessary to perform the task...
in that condition. Our finding that being allowed to gesture benefited participant descriptions more in the static than active content condition is interesting because this variable has not been used in past research. Future research could employ video clips with varying degrees of activity in their content, which provides the benefit of actual moving action rather than a still picture of inferred motion. Future research should continue to look at the influence of gestures on pauses as a measure of speech disfluency because we found no effects on this measure. These results were contrary to the findings of Graham and Heywood (1975) who found an increase in proportion of time spent pausing when participants were restricted from gesturing.

The limited number of gestures produced by participants in the gestures-allowed condition could be attributed to the fact that our picture description task did not require the use of significant mental imagery. Sassenberg and van der Meer (2010) suggested that gestures are used more often with activation of images in the mind (see also Wesp, Hesse, Keutmann, & Wheaton, 2001). The task of the current study only required participants to describe a visible picture; therefore, little mental imagery was needed to complete the task. Had participants been asked to describe a scene from their own imagination, rather than a scene presented to them, perhaps more gestures would have occurred. Mol, Krahmer, Maes, and Swerts (2009) found that participants used fewer gestures when giving descriptions to a machine than to a human, suggesting that the lack of gestures in the current study could have occurred because participants did not have an audience to whom they were communicating. Participants in our study were not told to describe the picture to the experimenter, and they generally directed their descriptions directly to the video camera. In future research, having an experimenter or confederate serve as the recipient of the information being conveyed by the speaker could increase the production of gestures during the picture description task.

Previous research on the impact of gestures on speech has yielded various results (e.g., Beattie & Coughlan, 1999; Frick-Horbury & Guttentag, 1998; Hostetter & Skirving, 2011; Pine et al., 2007; Ravizza, 2003). The present findings contributed to the literature on this relationship and suggested the need for more research in this area, with the ultimate goal of developing a sound theoretical model of how gesturing exerts an influence on speech fluency.

References

Author Note. Allison M. Douglass, Department of Psychology, University of Colorado, Colorado Springs; Lori E. James, Department of Psychology, University of Colorado, Colorado Springs.

This research was supported by a research award from the University of Colorado, Colorado Springs’ College of Letters, Arts, and Sciences to Allison Douglas, and served in partial fulfillment of her Honors Program requirements for the Psychology Department of the University of Colorado, Colorado Springs. Portions of this research were presented at the 2011 meeting of the Rocky Mountain Psychological Association, Salt Lake City, UT. The authors thank Austin Bowman, Justin Gentry, and Consuelo Garcia for research assistance.

Correspondence concerning this article should be addressed to Lori E. James, Department of Psychology, University of Colorado, Colorado Springs, 1420 Austin Bluffs Parkway, Colorado Springs, CO 80918. E-mail: ljames@uccs.edu