Journal of Experimental Psychology:
Human Perception and Performance
2013, Vol. 39, No. 5, 1496–1502

0096-1523/13/$12.00 DOI: 10.1037/a0033594

COMMENTARY

Does the Ventriloquist Illusion Assist Selective Listening?

Bradley N. Jack
Southern Cross University

Robert P. O'Shea
Southern Cross University and University of Otago

David Cottrell
James Cook University

Walter Ritter
Nathan Kline Institute for Psychiatric Research, Orangeburg, New York

Driver (1996) reported that the ventriloquist illusion can enhance selective listening of speech. Participants in his study listened to target and distractor words from a single loudspeaker while watching lip movements of the target words on a video monitor either above the loudspeaker or displaced to the left or right. He found that participants were more accurate in repeating the target words when the video was displaced from the loudspeaker than when the video was directly above the loudspeaker. Driver proposed that the ventriloquist illusion dragged the target sounds toward the location of the lip movements, freeing them from interference from the distractor words. However, successful attempts at replicating this finding are rare (we know of only three successful replications from 19 attempts). In five experiments, we found a weak advantage for selective listening from displaced lip movements only when there was a convincing ventriloquist illusion. We conclude that the ventriloquist illusion is necessary to confer the advantage for selective listening from displaced lip movements but that the phenomenon is a fleeting one at best.

Keywords: Driver (1996), selective listening, ventriloquist illusion, multisensory perception

Perception of speech depends on more than just the quality of auditory information. Visual information is also an important influence. For example, watching congruent lip movements, such as when at a cocktail party, helps selective listening of speech (Cherry, 1953; Sumby & Pollack, 1954), whereas watching incongruent lip movements, as in the McGurk effect, hinders hearing phonemes accurately (McGurk & MacDonald, 1976). Moreover, watching congruent lip movements in one location can hinder localizing speech sounds from another location—the ventriloquist illusion (Bertelson & de Gelder, 2004; Howard & Templeton, 1966).

We are interested in whether the ventriloquist illusion can assist selective listening. Driver (1996) found that viewing lip movements improves selective listening more when they are displaced from target speech sounds than when they are in the same place. In his Experiment 1, Driver played target and distractor speech sounds from a single loudspeaker located either to the left or right of an inactive loudspeaker. At the same time, he played a video of an actress making the lip movements of the target words on one of two monitors, each located just above one of the loudspeakers. On half the trials for each participant the active monitor was located above the active loudspeaker (the *same-side* condition; see Figure 1A); on the remaining trials, the active monitor was located above the inactive loudspeaker (the *displaced-side* condition; see Figure 1B). The participant's task was to shadow, that is to repeat, the target words.

Driver (1996) found that when the lip movements were on the same side as the target and distractor speech sounds, participants shadowed the target words with 59% accuracy. But when the lip movements were displaced from the speech sounds, participants shadowed them with 76% accuracy. We will call this improvement in selective listening *displacement-aided speech perception* (DASP). Driver proposed that when the lip movements are displaced, participants experience the target sounds as being pulled toward the active monitor—the ventriloquist illusion—and the distractor sounds as coming from their true source, the active

Bradley N. Jack, Discipline of Psychology, Cognitive Neuroscience Research Cluster, School of Health and Human Sciences, Southern Cross University, Coffs Harbour, Australia; Robert P. O'Shea, Discipline of Psychology, Cognitive Neuroscience Research Cluster, School of Health and Human Sciences, Southern Cross University and Department of Psychology, University of Otago, Dunedin, New Zealand; David Cottrell, Department of Psychology, James Cook University, Cairns, Australia; Walter Ritter, Cognitive Neurophysiology Laboratory, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY.

We thank Natasha Lindros, Christina Nguyen, and Eleanor Doyle, for assisting with data collection; Katie Peterson, for providing a fine speaking voice during the recording of our stimuli; Jon Driver, for helpful discussion; and three anonymous reviewers, for commenting helpfully on earlier versions of this article. Conceived the experiments: BJ, ROS, DC, WR. Designed the experiments: BJ, ROS, DC. Performed the experiments: BJ. Analyzed the data: BJ, ROS, DC. Contributed reagents/materials/analysis tools: DC. Wrote the article: BJ, ROS, DC, WR.

Correspondence concerning this article should be addressed to Bradley N. Jack, Discipline of Psychology, Southern Cross University, Hogbin Drive, Coffs Harbour NSW 2450, Australia. E-mail: bradley.jack@scu.edu.au

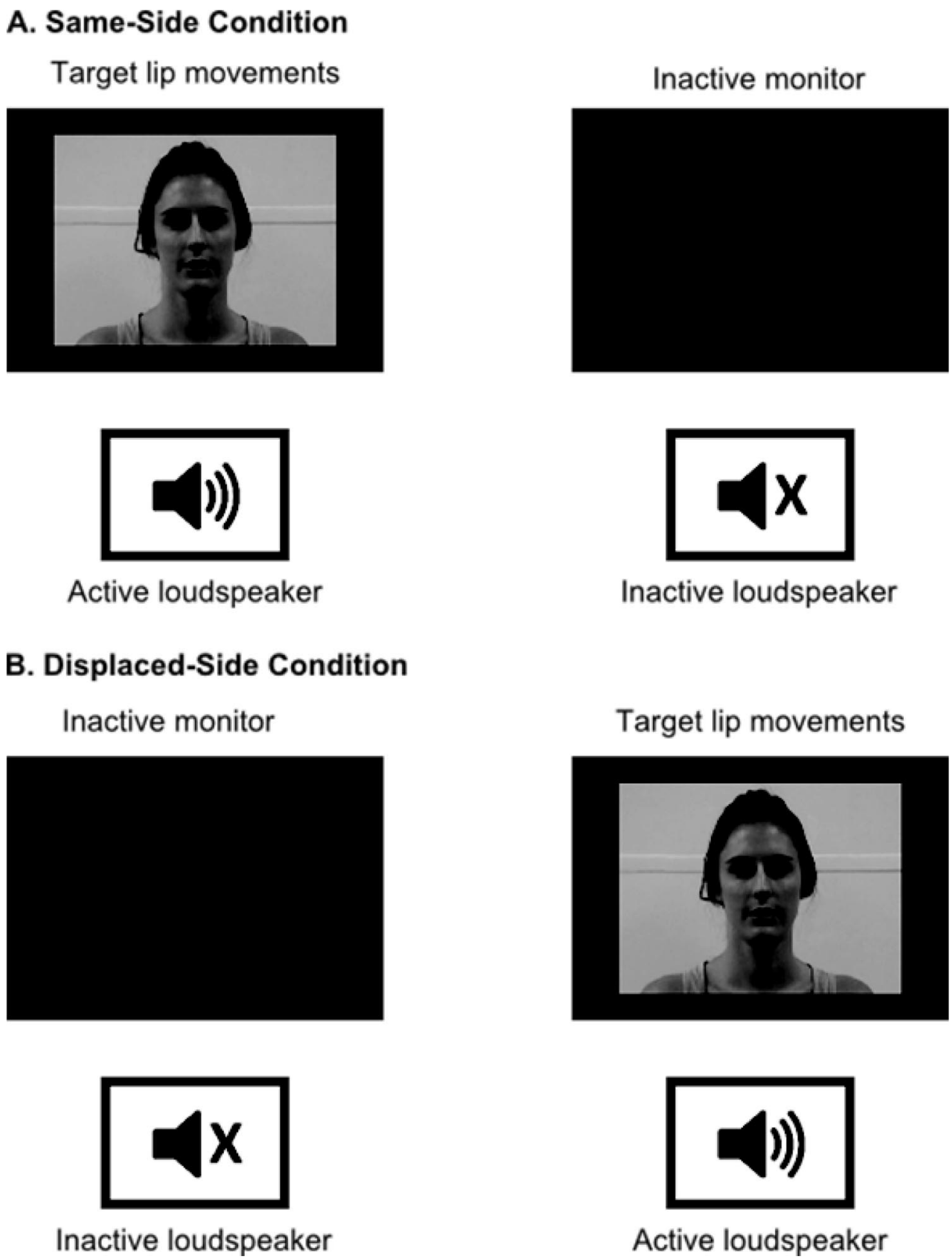

Figure 1. Schematic illustration of the apparatus used by Driver (1996) and in our Experiments 1–5. Target and distractor speech sounds were played from an active loudspeaker, located either to the left or right or an inactive loudspeaker. At the same time, lip movements matching the target words were played on either the monitor directly above the active loudspeaker (A. Same-Side Condition) or directly above the inactive loudspeaker (B. Displaced-Side Condition).

loudspeaker. We will call this last key element of the explanation *ventriloquism-aided speech perception* (VASP).

Driver (1966) also concluded that the ventriloquist illusion operates prior to the allocation of attention. To reach this conclusion, Driver conducted two additional experiments. In Experiment 2, he found that participants' viewing the lip movements of the distractor sounds reduced accuracy for the target sounds even though the target and distractor sounds came from different loudspeakers located on either side of a single monitor. That is, participants could not prevent audio-visual integration, even when it meant their accuracy suffered. In Experiment 3, Driver confirmed the importance of lip reading for shadowing when the target and distractor sounds came from essentially the same location below the monitor.

Driver's (1996) conclusion, that multisensory inputs are integrated before we attend to them, has largely been accepted. For example, a Web of Science cited reference search found almost 150 articles reporting Driver's conclusion. However, successful attempts at replicating this finding are rare. In fact, we know of only three successful replications (Heard & Webb, 2001; Leech, 1999, 2001) from 19 attempts (Brungart, Kordik, & Simpson, 2005; Fussell & Culling, 1999; Heard & Webb, 2001; Leech, 1997, 1999, 2001; Rudmann, McCarley, & Kramer, 2003). In the three experiments showing DASP, there was a strong ventriloquist illusion. In the experiments not showing DASP, four had a strong ventriloquist illusion (Leech, 1999, 2001), four had a weak ventriloquist illusion (Leech, 1997, 2001), and the rest did not provide information about the ventriloquist illusion (Brungart et al., 2005; Fussell & Culling, 1999; Leech, 1999, 2001; Rudmann et al., 2003). This is mostly consistent with VASP's being the explanation for DASP.

To attempt to replicate DASP, we spent considerable time pilot testing. We learned three important lessons:

1. Participants who knew the sound was coming from only one loudspeaker never experienced the ventriloquist illusion. This meant we could neither test ourselves nor experience DASP.

2. Participants who ran through the procedure more than once performed much better the second time, yielding ceiling effects. This meant we were forced to test each participant only once.

3. Participants whose attention was drawn to the location of the source of the sound tended not to experience the ventriloquist illusion or to show DASP. This meant we could not assess the ventriloquist illusion after every trial.

In Experiments 1–4, we assessed the ventriloquist illusion only once for each participant immediately after the shadowing task was completed. In Experiment 5, we manipulated the strength of the ventriloquist illusion by decreasing the size of the displacement in our apparatus to determine its effects on DASP. If the ventriloquist illusion is necessary for DASP, we should show DASP in experiments yielding a strong ventriloquist illusion but not in experiments yielding a weak ventriloquist illusion. If only displacement is necessary, then this should improve accuracy irrespective of the strength of the ventriloquist illusion.

Experiments 1–4

Method

Participants. One hundred and 32 [12; we provide Driver's (1996) parameters in square brackets where ours differ] naïve volunteers participated in our study (96 female, age $M = 31$ [22], $SD = 13$ years). All participated in only one experiment each. All reported having normal auditory and normal or corrected-to-normal visual acuity. All provided written informed consent and there was no reward or incentive offered to participate. The study was approved by the Southern Cross University Human Research Ethics Committee.

Apparatus. The apparatus varied between experiments. In Experiments 1 and 2, two Samsung monitors (SyncMasters, one a 2243BWPLUS, the other a 2233RZ, both with screen resolutions of 1280 × 800 pixels) were placed 102 cm apart (center-to-center), symmetrically to the left or right, on a plane 198 cm in front of the participant. The angle between them was 28.89 degrees. A Logitech Z320 loudspeaker was located below each monitor. The experiment was controlled using a PsyScope X script running on a MacPro computer. All participants were seated at a table and used a chin rest throughout testing.

In Experiment 3, a Sanyo PLC-WXU30 LCD data projector with a resolution of 1280 × 800 pixels projected the left-right visual stimuli 250 [102] cm apart (center-to-center), symmetrically to the left or right, on a screen 464 [198] cm in front of the participant. We chose these distances because they maintained Driver's (1996) original visual angle (28.89 degrees); thus, the apparatus was on a larger scale to that of Driver's but identical in spatial configuration. It was also conducted in a different room to that used in Experiments 1 and 2. An Athena Audition loudspeaker was located below each visual stimulus. Every other aspect of the apparatus was identical to Experiments 1 and 2.

In Experiment 4, we used the same apparatus described in Experiments 1 and 2, except that we replaced the loudspeakers with Sony MDR-XD100 stereo headphones. We used headphones because they abolish the effect of room cues on localizing auditory information.

Stimuli. To make the stimuli, we digitally video-recorded an actress who read 480 [336] two-syllable words from a computer display, at the location of the camera, presenting each word at the rate of one per second [one every 500 ms]. All the words were approximately matched for spoken word frequency in pairs [two-syllable, stress-initial nouns approximately matched for spoken word frequency and on number of letters] and all were digitally filtered to decrease any background noise (mostly 50 Hz hum). We cropped the video recording to show the head and shoulders [only the head] of the actress. These recordings were arranged into 80 [56] trials consisting of three target words and three distractor words. We used only the video of the target words [both videos; one after the other such that participants repeated words that had previously been distractors] and we adjusted both auditory tracks to ensure synchronization with the video.

In Experiments 1 and 2, the peak level of the verbal messages was 46–48 dB and the background noise was 40 dB [38 dB], giving a signal-to-noise ratio of 1.18 [1.24] at the participant's location. In Experiment 3, the peak level of the verbal messages was 67 dB and the background noise was 42 dB, giving a signal-to-noise ratio of 1.60. In Experiment 4, the peak level of the verbal messages was 54 dB and the background noise was 44 dB, giving a signal-to-noise ratio of 1.22. In Experiment 4, we also modified the audio files using the human head-related transfer function (HRTF; Panorama© WaveArts). This was done to create the stereophonic illusion that the speech sounds came from one or the other side of the display when listened to with headphones.

Procedure and design. Participants were tested individually. All read standard instructions (the same for each experiment) on the computer display: that two concurrent verbal messages, each comprising three words per trial, would be presented; that one of those messages would match the lip movements of the actress; and that the task was to repeat the words that matched the lip movements. Each participant was also informed that although the location of the video would alternate between the left and right side of the display, he or she should keep his or her head facing directly between them at all times, moving only the eyes to look at the lip movements.

A session began with four practice trials (these stimuli did not appear in the experiment proper and responses to these words were not analyzed). The following 80 trials were then run through once [twice; on the second run, words that had previously been targets were now distractors] in a different random order for each participant. All trials, including the practice trials, had the sound coming from only one loudspeaker (Experiments 1–3) or from one side (Experiment 4).

The true location of the speech sounds was counterbalanced over participants. At any one time during the session, only one side of the display was active. The left-right location of the video alternated over blocks of 10 [14] trials. The starting side of the video was randomized over participants. After the stimuli were delivered, participants attempted to shadow the target words. There was no set time for this, but participants responded within about 5 [4.5] seconds.

Participant responses were scored as either correct (reporting the target word) or incorrect (reporting a word other than the target word or not reporting a word). Scoring was not subject to the order of report for words within each trial. Close approximations (e.g., *sunsets* for *sunset*) were scored as incorrect. Next, the experimenter tallied the number of correct responses in the same-side and displaced-side conditions, and transformed these values into percentages.

After the experiment proper, which was completed in less than 15 minutes, participants were asked a set of gradually more-leading questions designed to elicit their thoughts about the purpose of the experiment (see Appendix A). In particular, we asked them where they thought the speech sounds were coming from. This gave us an index of whether or not participants experienced the ventriloquist illusion.

Results and Discussion

We assessed participants' shadowing accuracy in each experiment using a mixed analysis of variance (ANOVA) on the percent correct data. This had location of the speech sounds (left loudspeaker vs. right loudspeaker) as between-participants factors, and location of the speech sounds in comparison with the lip movements (same-side condition vs. displaced-side condition) as the within-participant factor.

The between-participants main effect (side of active loudspeaker) was not significant in any experiment, nor were there any significant interactions. The within-participant factor (same-side vs. displaced-side) was significant only in Experiment 1 (for more information, see Table 1).

Note that all experiments had a power greater than .999 to detect a statistically significant difference in shadowing accuracy of 17% (as found by Driver, 1996) with a standard deviation of 12%. Table 1 shows that when there is a strong ventriloquist illusion (Experiment 1) there is an advantage for selective listening from displaced lip movements, but that that advantage disappears when there is a weak ventriloquist illusion (Experiments 2–4). The critical experiments are 1 and 2. They were conducted in the same laboratory with the same apparatus, stimuli, and procedure, but at different times of the year, by different experimenters, and with some rearrangement of the laboratory (see later). Experiment 1 yielded DASP and a strong ventriloquist illusion; Experiments 2–4 yielded no evidence of DASP and a weak ventriloquist illusion. These findings are consistent with VASP's being the explanation for DASP.

Experiment 5

The relationship between the ventriloquist illusion and obtaining DASP in Experiments 1–4 provides correlational evidence that enhanced selective listening depends on the strength of the ventriloquist illusion, consistent with our hypothesis. Next, we decided to test this experimentally by attempting to manipulate the strength of the ventriloquist illusion.

The ventriloquist illusion is promoted by three factors.

1. The auditory and visual events should coincide temporally (Jack & Thurlow, 1973; Slutsky & Recanzone, 2001; Thurlow & Jack, 1973; Wallace et al., 2004). This is accomplished by synchronizing the audio-visual stimuli to within about 200 ms and by ensuring they are similar so that it is easy to imagine the two events' being related in the real world (i.e., lip movements and speech sounds are synonymous with face-to-face communication). In our experiments, the speech sounds and lip movements were carefully synchronized to within 40 ms so it is not an influence. Heard and Webb (2001), however, found DASP to occur with a 200 ms delay between auditory and visual information, but not with a 300 ms delay. This is consistent with what we know about the ventriloquist illusion.

2. The reliability of the information about the location of the speech sounds must not be too high (Alais & Burr, 2004; Thurlow & Jack, 1973). In an anechoic chamber, accuracy in locating the source of sound in the horizontal plane is about 1 degree (Mills, 1958), but in a poorly designed concert hall accuracy can be much worse because of misleading echoes from surfaces in the room (Hartman, 1983). In between Experiments 1 and 2 we changed the room by removing three filing cabinets and placing books on a bookshelf. It is possible the different room cues accounted for the differences in the ventrilo-

Table 1
Results From Four Experiments Giving Number of Participants, Means, and Standard Deviations of the Percentage of Shadowing Accuracy for the Same-Side and Displaced-Side Conditions, the Significance of the F Test for Separation, the Effect Size, and the Percentage of Participants Reporting the Ventriloquist Illusion (VI)

		Same-Side		Displaced-Side				
Experiment	*n*	Mean	*SD*	Mean	*SD*	*p*	η_p^2	VI
1. Replication 1	43	51	11	59	13	**<.001**	.41	86
2. Replication 2	16	55	8	56	9	.81	.004	25
3. Data projector	35	48	11	49	10	.53	.12	31
4. Headphones	38	69	9	68	9	.43	.02	26

quist illusion in these experiments. Experiment 3 was conducted in a different room that was much larger, which may have accounted for the weak ventriloquist illusion. Experiment 4 was conducted with headphones. It is possible that there was no ventriloquist illusion because participants failed to perceive the sounds as coming from outside the head (Blauert, 1996).

3. The horizontal separation between the location of the lip movements and the speech sounds should not exceed about 30 degrees (Jack & Thurlow, 1973; Jackson, 1953; Wallace et al., 2004; Witkin, Wapner, & Leventhal, 1952). Our separations in Experiments 1–4 and in Driver's (1996) Experiment 1, of 28.89 degrees, are therefore at the limit of displacements that yield a ventriloquist illusion in normal conditions. It is possible that our and others' failures to find DASP occurred because 28.89 degrees was too large a separation to yield the ventriloquist illusion in the rooms in which the experiments were conducted.

We decided to manipulate the strength of the ventriloquist illusion by decreasing the size of the displacement in our apparatus. The optimum displacement is one that is as large as possible while giving a convincing ventriloquist illusion. We conducted additional pilot studies and found a ventriloquist illusion on about 90% of trials with a 16.20 degrees separation, but only on 20% of trials with a 28.89 degrees separation. This is consistent with what we know about the ventriloquist illusion (Jack & Thurlow, 1973; Jackson, 1953; Wallace et al., 2004; Witkin et al., 1952). We tested these two displacements in Experiment 5.

Method

Participants. Forty-one naïve volunteers participated in our study (35 female, age $M = 30$, $SD = 12$ years). All reported having normal auditory and normal or corrected-to-normal visual acuity. All provided written informed consent and eligible students received course credit for participating. The study was approved by the James Cook University Human Research Ethics Committee.

Apparatus. We used the same apparatus described in Experiments 1 and 2, except that we used two Samsung monitors (SyncMaster 171S) and placed them either 56 cm apart (center-to-center), symmetrically to the left or right, or 102 cm apart. The angle between them was either 16.20 or 28.89 degrees, respectively. A Digitech XC5182 loudspeaker was located below each monitor. The experiment was also conducted in a different room from the others.

Stimuli. We used the same stimuli described in Experiments 1–3, except that the peak level of the verbal messages was 58–62 dB and the background noise was 51 dB, giving a signal-to-noise ratio of 1.18.

Procedure and design. Participants were randomly and unequally assigned to the two separations. There were 28 participants at the smaller separation and 13 at the larger, original separation. We tested fewer participants at the larger separation because we knew from pilot testing in the room that this would be unlikely to yield DASP.

At the smaller separation, there were two parts to the procedure, run in the same order for every participant:

1. The same shadowing task described in Experiments 1–4.

2. An assessment of the ventriloquist illusion. For each participant we presented all the trials again, in a similar fashion to the first task (although with a new random order), except that only the target sounds were presented. The participant's task was to press the left button on a response keypad when the sound appeared to come from the left loudspeaker and the right button when it appeared to come from the right loudspeaker.

At the larger separation, we conducted only the shadowing task described in Experiments 1–4 (we knew from the pilot testing that there would be essentially no ventriloquist illusion at this separation).

Results and Discussion

We assessed participants' shadowing accuracy using a mixed ANOVA on the percent correct data. We used the same between- and within-participant factors as before, except that we also included one extra between-participants factor: small separation versus large separation. Neither of the between-participants factors were significant, nor were there any significant interactions with the within-participant factor.

In particular, in the 16.20-degree-separation condition, shadowing accuracy was identical in the same-side and displaced-side conditions ($M = 65\%$; $SD = 11\%$). In the 28.89-degree-separation condition, shadowing accuracy was essentially identical in the same-side and displaced-side conditions ($M = 68\%$) although the standard deviations differed: 16% for the same-side condition and 11% for the displaced-side condition. There were no significant interactions. That is, we failed to find DASP, even with a smaller displacement.

We analyzed the data from the smaller separation more fully. We calculated an index of the strength of the ventriloquist illusion for each participant as measured in the second part of the procedure as: The number of displaced-side condition trials in which a participant reported the ventriloquist illusion divided by the number of same-side condition trials in which a participant reported the sound as coming from the active loudspeaker. These scores can range from 0, showing that a participant was more likely to report the sound as coming from the active loudspeaker than to report the ventriloquist illusion, through 1, showing that a participant was equally likely to report the ventriloquist illusion as to report the sound as coming from the active loudspeaker, to some high value, showing that a participant was more likely to report the ventriloquist illusion than to report the sound as coming from the active loudspeaker. The indices had a mean of 0.88 ($SD = 0.40$), suggesting that the ventriloquist illusion was rather weak.

Nevertheless, there was a significant correlation between the index of the strength of the ventriloquist illusion and the presence of DASP, $r(28) = .42$, $p = .027$. When we entered the index as a covariate in an analysis of covariance (ANCOVA) for the smaller separation, we found that shadowing accuracy in the displaced-side condition ($M = 64\%$, $SE = 3\%$) was significantly better than in the same-side condition ($M = 57\%$, $SE = 3\%$), $F(1, 25) = 5.07$, $p = .03$, $\eta_p^2 = .17$. As in our previous analyses, there was no significant interaction between side of presentation and displacement.

Overall, the results were mixed. Our hopes that DASP would be higher at the smaller separation than at the larger separation were dashed, possibly because the ventriloquist illusion was rather weak, even at the smaller displacement of 16.20 degrees. It is also possible that the ventriloquist illusion was even weaker in the first part of the experiment than in the second part, because there were two speech streams and only one video, which is different from the method we used to gather our data about the strength of the ventriloquist illusion and from the usual method for inducing it (e.g., Bertelson & de Gelder, 2004). However, at the smaller separation some participants gave reports in the second part of the procedure showing that they experienced the ventriloquist illusion, and those same participants' data showed DASP in the first part of the procedure. This is consistent with our proposal that the ventriloquist illusion is necessary for DASP.

General Discussion

We attempted to replicate the key conditions of Driver's (1996) Experiment 1. He found that performance in a selective listening task was better when target lip movements were played at a location displaced from target and distractor speech sounds. Our Experiments 1 and 5 are the first to provide evidence for Driver's conclusion that the ventriloquist illusion can enhance selective listening.

Although some of our results are consistent with VASP's being the explanation for DASP, we need to consider at least three other possible explanations for those results.

1. There may have been experimenter effects (Rosenthal, 1976). Three different experimenters conducted Experiments 1, 3, and 5; one of us (BNJ) conducted Experiments 2 and 4. Experimenter effects could have occurred via three means. First, the experimenter could have biased participants with instructions. Second, the experimenter may have subtly rewarded responses that fit with his or her hypothesis. Third, the experimenter may have made favorable mistakes when scoring the data. However, these all seem unlikely because we had identical, standard instructions that participants read from the display in each experiment, because each experimenter left the room during the experiment proper and because interrater reliabilities between experimenters from Experiments 1–4 and blind raters showed almost perfect agreement for scoring (Kappa values ranged between .96 and 1, with an overall mean of .98, $p < .0005$; Landis & Koch, 1977). It seems unlikely to us that experimenter effects occurred.

2. There were different participants in each experiment. This raises the prospect that sampling error is responsible. To calculate this, we first estimated the probability of any participant's showing DASP based on the proportion of all our participants who showed any DASP whatsoever; this was 0.58. Then using the binomial theorem we calculated the probability for Experiment 1 of finding at least 37 such participants by chance out of 43. This yielded $p =$ .0000063, which is rather unlikely. Moreover, when we analyzed the data from only those participants who experienced the ventriloquist illusion in Experiments 1–5 there was evidence of DASP only in Experiments 1 and 5. Nevertheless, we cannot rule out some other characteristic we failed to measure predicts DASP and that the number of participants possessing this characteristic differed in the experiments.

3. There were differences in the signal-to-noise ratio between experiments. It is possible that DASP occurs under a very narrow signal-to-noise ratio and that our failures to replicate were not within this range. However, this also seems unlikely because Experiments 1, 2, and 5 had almost identical signal-to-noise ratios but Experiments 1 and 5 yielded DASP whereas Experiment 2 did not. Moreover, Experiments 1–5 surround Driver's (1996) signal-to-noise ratio.

Our advantage for selective listening from displaced lip movements in Experiments 1 and 5 is less (at about 7%) than that reported by Driver (1996; at 17%), although more than reported by Leech (2001; at 4%). Unfortunately, we do not know the size of the advantage found by Heard and Webb (2001) or by Leech (1999). Our experiments indicate that to obtain a power of .95 one would need sample sizes ranging from 21 participants with a 7% effect to 32 with a 4% effect. It is possible that attempts to replicate the effect have suffered from inadequate power as Driver's initial report from 12 participants of a 17% effect might be at the larger end of the range of possible effects.

Our study also shows that DASP and the ventriloquist illusion are not as robust as other audio-visual illusions. For example, we had to use naïve participants to generate a convincing ventriloquist illusion; this is not applicable to other audio-visual phenomena, such as the McGurk effect (Liberman & Mattingly, 1985). Indeed, this is quite surprising, because the ventriloquist illusion is often characterized as being a bottom-up process (Bertelson & Aschersleben, 1998; Bertelson & de Gelder, 2004; Bertelson, Vroomen, de Gelder, & Driver, 2000), whereas our study (especially Lesson 3 from our initial pilot study) suggests that the ventriloquist illusion is affected by top-down influences. We confess that at present, we are unable to explain this curious anomaly, but we look forward to exploring this in the future.

Our results are also consistent with Driver's (1996) general conclusion: that the ventriloquist illusion operates prior to the allocation of attention. Recently, however, van Ee, van Boxtel, Parker, and Alais (2009) have found evidence that audio-visual integration is not automatic, but relies on voluntary attention. This is just one possible avenue for further investigation regarding our characterization of the ventriloquist illusion as being affected by top-down influences because it is possible that voluntary attention is the key factor distinguishing participants who showed DASP in our experiments and those who did not. Unfortunately, further work is needed to pick apart the mechanisms responsible for integrating sensory information from multiple modalities.

In conclusion, we have successfully replicated the results of Driver's (1996) Experiment 1 in our Experiments 1 and 5, showing that displacing the lip movements from the target speech sounds can yield increased accuracy in selective listening (i.e., DASP). Our study also shows that DASP does not occur without a convincing ventriloquist illusion, and that even when it does, it is weaker than Driver found. Participants show considerable variability in their sensitivity to the ventriloquist illusion even under identical observation conditions and this is a potential explanation for the apparent fragility of DASP and subsequent difficulties in replicating the effect. We look forward to our and others' future research examining Driver's curious finding.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5,* 482–489. doi:10.3758/BF03208826

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–178). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198524861.003.0007

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62,* 321–332. doi:10.3758/BF03205552

Blauert, J. (1996). *Spatial hearing*. Cambridge, MA: MIT Press.

Brungart, D. S., Kordik, A. J., & Simpson, B. D. (2005). Audio and visual cues in a two-talker divided attention speech-monitoring task. *Human Factors, 47,* 562–573. doi:10.1518/001872005774860023

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975–979. doi:10.1121/1.1907229

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381,* 66–68. doi:10.1038/381066a0

Fussell, C., & Culling, J. F. (1999, October). *The integration of auditory and visual attention in speech recognition.* Paper presented at the 1st International Multisensory Research Forum, Oxford, UK.

Hartmann, W. M. (1983). Localisation of sound in rooms. *Journal of the Acoustical Society of America, 74,* 1380–1391. doi:10.1121/1.390163

Heard, P. F., & Webb, P. (2001, August). *The ventriloquist effect occurs with auditory and visual delay.* Paper presented at the 24th European Conference on Visual Perception. Kusadasi, Turkey.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London, UK: Wiley.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills, 37,* 967–979. doi:10.2466/pms.1973.37.3.967

Jackson, C. V. (1953). Visual factors in auditory localization. *The Quarterly Journal of Experimental Psychology, 5,* 52–65. doi:10.1080/17470215308416626

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. doi:10.2307/2529310

Leech, S. M. (1997). *The ventriloquist effect, illusory and real separation of auditory signals, and selective listening*. (Unpublished MSc thesis). University of Sussex, UK.

Leech, S. M. (1999, October). *The integration of auditory and visual attention in speech recognition.* Paper presented at the 1st International Multisensory Research Forum. Oxford, UK.

Leech, S. M. (2001). *The effect on audiovisual speech perception of auditory and visual source separation*. (Unpublished PhD thesis). University of Sussex, UK.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36. doi:10.1016/0010-0277(85)90021-6

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748. doi:10.1038/264746a0

Mills, A. (1958). On the minimum audible angle. *Journal of the Acoustical Society of America, 30,* 237–246. doi:10.1121/1.1909553

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York, NY: Wiley.

Rudmann, D. S., McCarley, J. S., & Kramer, A. F. (2003). Bimodal displays improve speech comprehension in environments with multiple speakers. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45,* 329–336. doi:10.1518/hfes.45.2.329.27237

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport, 12,* 7–10. doi:10.1097/00001756-200101220-00009

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215. doi:10.1121/1.1907309

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the "ventriloquism effect". *Perceptual and Motor Skills, 36,* 1171–1184. doi:10.2466/pms.1973.36.3c.1171

van Ee, R., van Boxtel, J. J., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Neuroscience, 29,* 11641–11649. doi:10.1523/JNEUROSCI.0873-09.2009

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research, 158,* 252–258. doi:10.1007/s00221-004-1899-9

Witkin, H. A., Wapner, S., & Leventhal, T. (1952). Sound localisation with conflicting visual and auditory cues. *Journal of Experimental Psychology, 43,* 58–67. doi:10.1037/h0055889

Appendix A

Post-Experiment Questionnaire

After Experiments 1–4 concluded, each experimenter asked each participant the following questions and recorded the answers verbatim.

1. What do you think was the purpose of this experiment?
2. Did you notice anything unusual?
3. Did you notice anything unusual about the sound?
4. Where do you think the sound came from?
5. Did you think the sound was coming from both of the speakers?
6. Did it ever sound as if different words were coming from different speakers?
7. If so, which ones?
8. How difficult did you find identifying the words that matched the lips when the sound and face were on the same side?
9. How difficult did you find identifying the words that matched the lips when the sound and face were on the opposite side to each other?
10. Did you realize all sounds were coming from only one speaker?
11. In any case, which speaker do you think it was?

Received January 21, 2013
Revision received April 9, 2013
Accepted April 12, 2013 ■