

# Speaking through the body

## *Do people associate the body movements of politicians with their speech?*

Markus Koppensteiner, *University of Vienna*

Greg Siegle, *University of Pittsburgh and Netherlands Institute for Advanced Study*

**ABSTRACT.** When people speak, they gesture. However, is the audience watching a speaker who is sensitive to this link? We translated the body movements of politicians into stick-figure animations and separated the visual from the audio channel. We then asked participants to match a selection of five audio tracks (including the correct one) with the stick-figure animations. The participants made correct decisions in 65% of all cases (chance level of 20%). Matching voices with animations was less difficult when politicians showed expansive movements and spoke with a loud voice. Thus, people are sensitive to the link between motion cues and vocal cues, and this link appears to become even more apparent when a speaker shows expressive behaviors. Future work will have to refine and validate the methods applied and investigate how mismatches between communication channels affect the impressions that people form of politicians.

**Key words:** Nonverbal communication, multimodal signals, bodily expressiveness, vocal expressiveness, gesture, impression formation, stick figure animation, politics

When people speak, they gesticulate. Apart from the movements produced by the mouth, the movements of the eyes, eyebrows, head, hands, and arms accompany speech and illustrate and emphasize what is being said.<sup>1,2,3,4,5</sup> Research investigating the role of body motion in human communication suggests that gestures and body movements can convey meaning independently of auditory information (e.g., a head nod as a gesture of approval). However, usually the two communication channels are strongly intertwined, making it difficult to interpret a specific gesture when no verbal information is present (e.g., illustrating with the hands how an object looks). Microscopic analyses of movement-speech coordination have revealed that gesturing is linked to verbal content as well as rhythmically aligned with speech on the level of syllables and phonemes and other nonverbal information.<sup>6,7</sup> However, sometimes gesturing is not strictly in sync with speech but rather precedes it,

thereby helping an interaction partner or an audience watching someone giving a speech anticipate what comes next.<sup>8</sup>

The observation that gestures and speech are intertwined gave rise to the idea that they share a common psychological structure.<sup>8,9</sup> Some researchers suggest that cognition — and language production is considered as cognition — is “grounded” in bodily action.<sup>10</sup> Such claims are supported by neurocognitive studies. These studies provide evidence that speech and gesturing are linked together by common brain activity.<sup>11,12</sup> Moreover, the synchrony between gesturing and speech appears to enrich communication and facilitate cognitive processes.<sup>13</sup> For instance, language comprehension is enhanced when gesturing is synchronized with information from the auditory channel. Also, gesturing appears to help people retrieve words from their mental lexicon. Telling people not to move their hands during an explanation task impairs information recall.<sup>14,15,16,17</sup>

Apart from facilitating language-related cognition and complementing each other when people communicate, auditory and visual cues also have an affective or relational component that guides interpersonal communication.<sup>1,18,19,20</sup> Acoustic features of speech,

**doi:** 10.1017/pls.2017.23

Correspondence: Markus Koppensteiner, NIAS/KNAW, Korte Spinhuissteeg 3, 1012 CG Amsterdam, The Netherlands. Email: [markus-koppensteiner@gmx.net](mailto:markus-koppensteiner@gmx.net)

body postures, and body motion can be powerful communicators of emotional states and affect impression formation.<sup>21,22,23,24,25,26</sup> This also transfers to the domain of politics. For instance, simple stick-figure animations representing the body movements of politicians preserve enough interpersonal information to enable attributions of personality traits or judgments of a speaker's health.<sup>27,28</sup> Experiments using manipulated voices, on the other hand, provide evidence that vocal cues affect attributions of leadership qualities as well as voting behavior.<sup>29,30</sup> Studies focusing on more than one nonverbal channel have shown that people are able to recognize statements of agreement or disagreement in political debates on the basis of low-level auditory and motion cues.<sup>31</sup> Moreover, apart from other nonverbal cues, the tonal elements and gestures that politicians produce during debates can serve as predictors of people's reactions on social media.<sup>32,33</sup>

In general, people appear to be capable of making use of information from a variety of communication channels, such as appearance features, facial expressions, gestures, and verbal content, when they interact with and form impressions of their social environment.<sup>21,27,31</sup> Thus, political candidates and officeholders entering the public stage are judged by their verbal and nonverbal communication skills.<sup>34</sup> Questions arise regarding how information from different modalities is weighted by perceivers, what kind of information grabs their attention, and what kind of information combines to make a "message" more salient. Research on multisensory processing shows that people's capacity to pay attention to multiple messages is limited.<sup>35,36</sup> However, it depends on the type of information that is presented simultaneously. For instance, when there is information redundancy between different modalities (e.g., audio and video channel) a message is better remembered.<sup>36</sup> Also, many events in our environment generate stimuli of several modalities (e.g., a moving car produces noise), and this makes it more likely that they create one perceptual unit.<sup>37,38</sup>

People form expectations of what kind of stimuli go together, and it sometimes only becomes apparent that they are a composite when these expectations are violated (see Discussion). The audio-visual link between mouth movements and speech is particularly strong, and seeing the movements of the mouth facilitates the comprehension of spoken words.<sup>39</sup> Given that speech and body gestures are also very often in sync, they appear to form a perceptual unit. Such an idea is supported by empirical findings. In the aforementioned

experiment on disagreement or agreement displays during a debate, people were better at decoding nonverbal messages when they had access to both body motion and corresponding auditory information.<sup>31</sup>

Researchers in the field of animal communication have observed that social signals are often communicated through different sensory modalities.<sup>40,41</sup> It is assumed that this serves two main purposes. First, sending redundant signals on multiple channels enhances the probability that the message will be transferred successfully. In this case, redundancy is considered as a kind of backup that makes communication more reliable. Second, multiple signals can convey different messages simultaneously. Regardless of which hypothesis applies (this might vary depending on the situation), multi-component signals seem to improve the recognition of stimuli and important social information.<sup>41</sup>

As biologists consider communication not only as information transfer but also as a means to gain social influence,<sup>42</sup> multimodal signals might have emerged over evolutionary history as a result of signaler and perceiver roles.<sup>43</sup> On the signal sender's part (e.g., the political candidate), multiple signals might help to make a message clearer, while on the perceiver's part (e.g., the potential voter), paying attention to multiple indicators might help avoid being manipulated. In summary, despite creating a cognitive load, multimodal signals appear to have a clear function in human communication.

Interrelated (or redundant) signals or cues that are communicated through different modalities might be the result of an evolutionary arms race between signal sender and signal receiver. Moreover, some human communication channels appear to be strongly interconnected and therefore form a kind of perceptual unit. Given that vocal utterances are often accompanied by gestures, this might be particularly true for body motion and vocal cues. Consequently, it is a plausible conclusion that people can associate the voices of politicians with the body movements of the same politicians.

To support this hypothesis, we performed an experiment for which we used short excerpts of speeches held in the German parliament. The body motion of the speakers was converted into stick figures. The audio tracks of the speakers' voices were used in an unmodified way. Overall, behaviors and utterances were from an ecologically valid source. The experimental procedure applied was different from a "classical" rating experiment, during which stimuli are judged on a set of verbal items. Instead, people were asked to assign speech segments to the movements of the speakers.

This study can be considered a first step in using the stick-figure method to investigate the relationship between vocal utterances and body movements.

The aims of the current work can be summarized as follows: First, we tested whether people are capable of correctly identifying those speech segments (among a selection of speech segments) that correspond with the speakers' body movements (represented as stick-figure animations). Thus, we intended to demonstrate that the observers of political speeches are sensitive to the language-gesture link. Second, we aimed to determine whether the level of difficulty for identifying the correct speech segment varies from stimulus to stimulus. We assumed that speakers with expressive styles of presenting themselves render it less difficult to make the correct choices. For this reason, we tested whether more expansive movements and louder voices — cues that have been shown to affect perceptions of dominance<sup>44,45</sup> — make it easier to make the correct associations. Third, as an addition to point two, we also examined whether people associate expressive behaviors with expressive vocal utterances (i.e., louder voices) when they make incorrect choices (i.e., when they did not assign the correct speech segment to a stick-figure animation).

## Method

### *Participants*

A total of 64 students (Caucasian; 33 females and 31 males; *M* age = 23.08 years, *SD* age = 4.28) were recruited to take part in an experiment. Recruitment took place at the Faculty of Life Sciences of the University of Vienna. Participants were students from different subfields of biology (zoology, ecology, etc.) and received a financial compensation of €5 for taking part in the experiment.

### *Stimulus preparation*

Stimuli presented during the experiment have already been used in previous work.<sup>27</sup> Source material included 60 speeches given in the German parliament (30 male and 30 female speakers). These speeches were taken from three parliamentary sessions (November 29–30 and December 14, 2012) using a random number generator. Deviations from random selection were necessary to reach equal numbers of male and female speakers and nearly equal numbers of different party members (i.e., members of Bündnis 90/Die Grünen, Christian Democratic Union/Christian Social Union, Die Linke,

Free Democratic Party, and Social Democratic Party). Starting at randomly selected positions, short video excerpts (15 seconds) were extracted from each of the speeches. Sometimes the first excerpt was not usable because a member of the parliament passed the camera (i.e., members sometimes walked around and obscured the speaker for a short moment). In addition, we also dismissed sequences during which the speakers were holding an object in their hands (e.g., a sheet of paper) or reading aloud from a piece of text. In these cases, the random search was restarted to select another sequence. The length of the sequences was a compromise between workload for encoding (i.e., the encoding procedure to create stick figures is very time-consuming) and enough variation in body motion. Also, it is within the length of sequences that are commonly used in impression formation research.<sup>18</sup>

In the next step, the body movements of the speakers were converted into animated stick figures. To create these stick-figure animations, the custom-made program *SpeechAnalyzer* was used. With this software, it was possible to run through a video frame by frame and capture motion by positioning landmarks (i.e., dots placed on the computer screen with the mouse) on different body parts. Starting at the first frame of each video, the landmarks were positioned on the speaker's forehead, hollow of the throat, chin, ears, shoulders, elbows, and hands; the corners of the desk; and the center of gravity (see Figure 1). By moving through the video step by step and rearranging the landmarks with the mouse and the support of automatic tracking software routines (i.e., optical flow), position shifts of the body parts were recorded. More precisely, the body movements of the politicians were stored as a time series of two-dimensional coordinates (all landmarks from frame *t* to *t* + length of video). Drawing lines between coordinates of each encoded frame gave a succession of stick figures,<sup>46</sup> which were turned into videos representing the body movements of the speakers in an abstract manner (see Figure 1). Because capturing body movements was a time-consuming procedure, we only used every third frame. To arrive at the same frame rate as the original video, missing frames were filled in by linear interpolation (i.e., interpolation between corresponding coordinates of successive frames).

Audio tracks were extracted from the 15-second-long sequences and used for the experiment in an unmodified way. As the contributions of the speakers were randomly selected, the speeches touched on different topics.



**Figure 1.** Transformation of a politician's body movements into stick figures (top-left image gives the names of the landmarks used during encoding process).

The speakers discussed — just to give a few examples — the national budget, health care, fight against corruption, agriculture, etc. However, in the brief excerpts we extracted, the topics of the full length speeches were hardly recognizable (see the online supplement for a list of the contents of the audio tracks used).

To sum up, speeches were decomposed into their auditory information and into the motion information contained in the stick-figure videos.

### *Procedure*

For the experiment, participants were brought to the laboratory, where they received instructions on how to operate the software we used for the experiments. The software was easy to handle and guided the participants through the whole procedure. The participants performed the tasks on their own without the help of an experimenter.

On the left-hand side of the software interface, the stick-figure videos were presented (i.e., video window). On the right-hand side, there were five radio buttons that were named “Tonspur” (i.e., audio track). The radio buttons were numbered consecutively, 1 to 5. When the experiment was started by clicking on the “start” button, a stick-figure video was randomly selected from the 60 videos available and played in the video window. The program also randomly selected four audio tracks from the 60 audio tracks available, as well as the “correct” audio track (i.e., the audio track belonging to the stick-figure video that was played). The five audio tracks were randomly assigned to the five radio buttons. Clicking on one of the radio buttons started an audio track and a stick-figure video. Clicking on another radio button started another audio track and restarted the stick-figure video. After choosing the best-fitting audio track, the participants pressed the “next” button to start the next round of the experiment. One experimental session consisted of 15 rounds. During the experiment, participants were wearing headphones (AKG K 272 HD). The volume of the sounds during the experiment was kept constant.

#### *Estimates of bodily and vocal expressiveness*

To obtain a simple measure of stimulus bodily activity, which was intended to represent overall “bodily expressiveness,” we made use of the coordinate data recorded during the encoding of the speeches. It has already been shown in previous work that simple yet still informative measures (i.e., measures of overall expansiveness, velocity etc.) can be created on the basis of only four landmark coordinates (i.e., left and right hand, shoulder, forehead).<sup>47</sup> In this study, we did the same and summed the coordinates of these four landmarks for each encoded frame and extracted the distances between successive maximum “stretches.” More precisely, we defined a reference point (the first frame) and measured the distances from this reference point until they reached a maximum. The point at this maximum then served as a new reference point, and the procedure started again.

For instance, if speakers raised their arms (see the succession of images in Figure 1), the maximal distance was reached just before the arms came down again. The coordinates at this maximal distance then served as new reference point (i.e., the second reference point because the starting frame was first reference point) until another a maximum was reached (i.e., the third reference point). Thus, one cycle of raising and lowering the arms

provided two distances: the distance between the first frame and the second reference point and the distance between the second reference point and the third reference point. Doing this for a whole sequence of movements resulted in a time series of amplitudes capturing the overall expansiveness of motion.<sup>47</sup> Calculating the average amplitude (in pixels) for each speaker served as a rough estimate of the speakers’ overall “bodily expressiveness.” The higher the value of this estimate, the more expansive a speaker’s movements were.

Using custom MATLAB routines, we created an estimate of “vocal expressiveness” by extracting the volume of the voices (i.e., standard deviation of the sound signal). Volume was determined for chunks of one second and then an average across all units (i.e., across the 15 seconds of the sequences) was calculated. Thus, vocal expressiveness was simply a measure of how loud the voice of a speaker was. In addition, we calculated and examined five other vocal parameters, including speech rate, energy, pitch, formant, and the mel-frequency cepstrum coefficients, using MATLAB routines given in Ma.<sup>48</sup> We conducted exploratory analyses using these parameters (i.e., correlations between the parameters and the recognition rate). This did not yield any noteworthy results. Data processing and statistical analyses were carried out in the program R.<sup>49</sup>

## Results

Participants went through subsets of 15 stick-figure stimuli drawn from the 60 stick-figure animations that were available. During each of the 15 rounds of the experiment, they had five audio tracks to choose from. Thus, the probability of choosing the “correct voice” (i.e., the audio track belonging to the stick figure) by chance is 1 in 5 per round. This means one can expect three correct answers (i.e.,  $1/5 * 15$ ) for one experimental session if choices are made randomly. The number of correct answers per rater ranged from 2 to 15, with a median of 10. Overall, there were 960 rounds. The participants selected the correct audio track in 625 cases and the wrong audio track in 335 cases. This equals an average probability of success of 0.65 (95% CI [0.62, 0.68]).

In a second step, the number of correct matches for each stick figure was determined. On average, each stick figure was used as a stimulus 16 times (with a range from 15 to 17) throughout all experimental sessions. On the basis of how often a stick figure was used as a stimulus and the number of correct classifications,

a “recognition rate” was calculated (i.e., the relative frequency of how often each stick figure was correctly matched with its corresponding voice). The recognition rate ranged from 0 to 1. As it turned out, there were stick figures (three of them) that were never matched with the correct voice as well as stick figures (four of them) that were matched with the correct voice each time.

In a next step, we determined to what extent “bodily expressiveness” (i.e., overall amplitude of body movements) and “vocal expressiveness” (i.e., loudness of voice) were related to the number of correct identifications per stimulus (i.e., recognition rate). Results are shown in Table 1. Regression estimates ( $\beta$ -weights), the coefficients of bivariate correlations, as well as the relative weights, which give the independent contribution of the regression estimates to the regression model, are presented.<sup>50</sup> These estimates reveal that participants experienced less difficulty in assigning the correct speech segment to a stick-figure animation with expansive movements than to a stick-figure displaying less expansive movements. A similar effect was found for sound volume. The louder the voice, the more likely it was that the participants made a correct decision. The correlation between the predictors of the regression model was  $r(58) = 0.51$ , 95% CI [0.29, 0.68]. In other words, the more expansive the body movements of the speaker, the louder the voice. To sum up, there was a link between our estimates of bodily and vocal expressiveness. Also, participants appeared to experience less difficulty making correct choices when they encountered speakers (i.e., their stick-figure versions) who showed expansive body movements accompanied by a loud voice.

We also tested whether people tended to relate expressive body movements with high sound volume when failing to find the corresponding voices. To accomplish this, we filtered all the cases in which incorrect choices were made. Then we calculated a mixed model with the volume of the selected audio track as the independent variable, the expansiveness of motion as the dependent variable, and the raters as random factor. Because we used  $z$ -transformed data and one predictor only for our model, the  $\beta$ -weight of the regression can be interpreted in a similar manner as a correlation coefficient. The procedure yielded a coefficient of  $\beta = 0.17$ ,  $t = 3.079$  on the basis of 335 observations and 62 raters (the sample size is lower in this case because two raters gave 15 correct answers). Although this shows that there was a tendency to assign louder voices to “louder” movements, the effect size ( $\beta$ -weight) was not very pronounced.

**Table 1. Results for amplitude of body motion and volume of voices with recognition rate.**

	Reg. Rate	Amplitude	Volume
$M$ ( $SD$ )	0.65 (0.25)	19.9 (9.2)	0.05 (0.01)
$\beta$ -weight ( $SE$ )	—	0.40 (0.13)*	0.19 (0.13)
$RW$	—	0.18	0.09
$r$ [95% CI]	—	0.51* [0.29, 0.68]	0.40* [0.16, 0.59]

Notes: Reg. Rate = recognition rate (relative frequency of correctly identified voice motion pairings);  $RW$  = relative weight (gives explained variance of single predictor in regression);  $R^2 = 0.27$ ;  $df = 57$  for regression;  $df = 58$  for correlation.

\*  $p < 0.05$ .

## Discussion

Politicians’ utterances are accompanied by gestures. These gestures can be linked to the verbal content of what is being said (e.g., illustrating or even replacing verbal content), or they can be in accordance with the nonverbal information (i.e., prosody) conveyed by human speech.<sup>1,2,3,4,5</sup> In this work, we show that observers are able to associate characteristics of vocal utterances with characteristics of body motion. More precisely, we found that people are quite successful in correctly assigning the audio recordings of a speech (among a selection of audio recordings of speeches) to the corresponding body movements of the politicians (i.e., displayed as stick figures) giving the speeches.

For our experiment, we used stick-figure animations and audio recordings that were based on speeches given in the German parliament. Unlike other studies, we did not ask actors to perform behaviors that might only reflect the actors’ ideas of how politicians speak and gesture. Thus, the stimuli were from an ecologically valid source. In addition, the stick figures provide parsimonious representations of the speakers’ body movements that are free from confounding variables. More precisely, because the animations mainly capture information about the flow, the amplitude and the quantity of motion created by the body, head, and arms of the speakers, they help isolate information from a specific source (i.e., motion cues) while removing information from other nonverbal sources such as clothing and facial expressions.

Although the stick-figure method makes it possible to control for confounding variables, it also has drawbacks. Details such as finger movements and hand positions are not captured by this method, and it only gives a two-dimensional representation of body motion. It has already been shown that for some personality ratings,

there is a correspondence between the original movements of the speakers and their stick-figure versions.<sup>27</sup> Nevertheless, future work requires additional experiments with different types of stimuli in order to determine whether the stylization procedure had an influence on the results obtained. Despite the limitations of the method, the findings show that the stick figures preserved enough information to enable people to make correct associations quite often. Apparently, the changes in direction of the speakers' arms, head, and overall body movements (i.e., lifting and lowering and moving from the left to the right and vice versa) have something in common with auditory features that are embedded in the voices of the speakers.

We asked people to match body motion with auditory information by having them find the correct audio track among a random selection of incorrect choices. The order of presentation as well as the selection of stick-figure animations was also randomized. Consequently, although there were overlaps, none of the participants encountered the same selection of stimuli. In this first step, we regard such a setup as useful, as it gives insight into people's general ability to relate motion to corresponding vocal information. However, a drawback of random selection is that it makes it impossible to compare ratings of different participants. In follow-up work, we will have to present the same sets of stimuli to all participants in order to analyze individual differences in making correct associations and to determine interrater agreement. Knowing that expressive behaviors make the task of finding the correct match easier helps to create sets of stimuli that allow more systematic investigations (e.g., combining expressive voices with nonexpressive ones) than in the current study.

Further work is also needed to conduct a refined analysis of behavioral and vocal cues and the ways in which they are associated. The current method does not allow for a clear identification of the basis of those associations. There is a need to disentangle nonverbal vocal cues from language content and to examine how motion cues are related with prosodic features by applying methods used by other researchers<sup>6,7,51</sup> or using stimuli that will be manipulated in systematic ways (e.g., changing pitch or amplitude of gestures). Such in-depth analysis will not only give better insight into the specific elements that enable intermodal associations, it will also show whether the language-body motion link is more pronounced for some personality types or for people who are emotionally aroused.

The link between vocal and auditory information appeared to be differently pronounced for different stimuli and this made the task of assigning the correct audio tracks sometimes more and sometimes less difficult. We found that such differences were partly attributable to variations in bodily and vocal expressiveness. It was easier for the participants to identify the correct matching for speakers with expansive body movements (indexed by high average amplitude of overall bodily activity) and loud voices. We also found that expansive body motion often goes together with a high volume in speaking. These findings highlight the multimodality of human communication.<sup>21</sup> As perceptions of dominance are related to expansiveness in motion,<sup>45</sup> one could speculate that the connection between body motion and vocal features is easier to detect when speakers are emotionally involved or display dominance. Moreover, different personalities might differ in the expressiveness of their performances. To provide evidence for such speculations, further studies are required.

People experience difficulty executing parallel processing of messages that are presented simultaneously.<sup>35,36</sup> However, when information from different modalities form one common message, because they complement each other, such a message is better remembered.<sup>36</sup> In addition, researchers in animal communication assume and provide evidence that sending redundant signals increases the likelihood that a message is transferred successfully.<sup>41</sup> Politicians may indeed combine — even without being aware of it — expressive gesturing with vocal expressiveness in order to attach more importance to what they intend to broadcast. That the participants in the experiment had less difficulty making correct choices for stimuli with expansive movements and loud voices supports such an assumption. It is also in line with previous work investigating the role of different modalities in human communication. For instance, when rating politicians on the personality dimension of extraversion, people appear to be influenced by motion cues (represented as a stick figure) and vocal cues.<sup>27</sup> Moreover, people are better at recognizing statements of disagreement and agreement in nonverbal behaviors of politicians debating when both auditory and motion cues are available<sup>31</sup> — and they recognize emotions in bodily movements more easily when combined with consistent auditory information.<sup>52</sup> In summary, audiences (or signal perceivers) appear to be able to integrate information from both vocalic and motion stimulus channels when making social judgments. Politicians

(or signal senders), on the other hand, appear to make use of both channels to place more emphasis on their messages.

Presenting oneself on the public stage is a demanding task, and although many politicians receive coaching by professional communicators, audiences may sometimes perceive information from different communication channels not to be in harmony. People often expect their social environment to behave in a certain way and react strongly when their expectations are violated.<sup>53</sup> Behaviors that come across as atypical and deviant frequently lead to negative evaluations. Disharmonies between different nonverbal communication channels can, for instance, make leaders appear less charismatic.<sup>54</sup> It seems that leadership displays creating low expectancy violations are composed of behaviors that produce no incongruences between the different verbal and nonverbal levels and that are tailored to the situational context. Indeed, experiments using eye tracking have shown that inappropriate displays receive more attention from observers and are more negatively evaluated than appropriate displays.<sup>55</sup>

With the stimuli at hand, research in this domain can be extended. Because the stick figures are composed of coordinate data, they can be manipulated in a systematic manner (e.g., by making movements smaller) to further examine expectancy violations. Stimuli altered in this way can be then used in a rating experiment to deepen our understanding of which way inconsistencies in motion and voice affect perceptions of social qualities such as authenticity or trustworthiness. Such manipulations have the potential to extend the methodological repertoire to test whether cues from different modalities conveying a specific social quality (e.g., voices and movements conveying dominance) add up to give an even stronger impression of that specific quality.

People are sensitive to the link between motion and auditory cues when watching politicians giving a speech. Moreover, expressive body motion (i.e., high overall amplitude in body movements) and expressive vocal performances (i.e., louder voices) often appear to go together, and this makes it easier to perceive the link between both modalities. Follow-up work should elaborate on this and investigate which features people use to perceive similar patterns in vocal and motion information. Further tests must also be conducted to determine the ways in which inconsistent information affects people's assessments of politicians. This may help clarify whether strong intermodal connections (i.e.,

redundancy between speech and body motion) makes a message even more convincing.

## Acknowledgements

Markus Koppensteiner thanks Pia Stephan and Johannes Jäschke for helping collect the data. We thank Nadia Latif and Pia Stephan for proofreading the manuscript. This research is based on work funded by the Austrian Science Fund (FWF): P 25262-G16, the Netherlands Institute for Advanced Studies (NIAS/KNAW), the EURIAS Fellowship Programme, and the European Commission (Marie-Sklodowska-Curie Actions — COFUND Programme — FP7).

## References

1. G. Bente and N. C. Krämer, "Psychologische Aspekte bei der Implementierung und Evaluation nonverbal agierender Interface-Agenten," in *Mensch & Computer*, H. Oberquelle, R. Oppermann, and J. Krause, eds. (Berlin: Springer, 2001), pp. 275–285.
2. P. Bull and G. Connelly, "Body movement and emphasis in speech," *Journal of Nonverbal Behavior*, 1985, 9(3): 169–187.
3. P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Semiotica*, 1969, 1(1): 49–98.
4. R. M. Krauss, Y. Chen, and P. Chawla, "Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?" *Advances in Experimental Social Psychology*, 1996, 28: 389–450.
5. P. Wagner, Z. Malisz, and S. Kopp, "Gesture and speech in interaction: An overview," *Speech Communication*, 2014, 57: 209–232.
6. W. S. Condon and W. D. Ogston, "Sound film analysis of normal and pathological behavior patterns," *Journal of Nervous and Mental Disease*, 1966, 143(4): 338–347.
7. D. P. Loehr, *Gesture and intonation*, Unpublished doctoral dissertation. Department of Linguistics, Georgetown University, Washington, DC 2004.
8. A. Kendon, "Gesticulation and speech: Two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, Mary Ritchie Key, ed. (New York: Mouton, 1980), pp. 207–227.
9. D. McNeill, "So you think gestures are nonverbal?" *Psychological Review*, 1985, 92(3): 350.



10. L. W. Barsalou, "Grounded cognition," *Annual Review of Psychology*, 2008, 59: 617–645.
11. M. Gentilucci and R. D. Volta, "Spoken language and arm gestures are controlled by the same motor control system," *Quarterly Journal of Experimental Psychology*, 2008, 61(6): 944–957.
12. R. M. Willems and P. Hagoort, "Neural evidence for the interplay between language, gesture, and action: A review," *Brain and Language*, 2007, 101(3): 278–289.
13. S. Goldin-Meadow, "The role of gesture in communication and thinking," *Trends in Cognitive Science*, 1999, 3(11): 419–429.
14. P. Bernardis and M. Gentilucci, "Speech and gesture share the same communication system," *Neuropsychologia*, 2006, 44(2): 178–190.
15. S. Goldin-Meadow, H. Nusbaum, S. D. Kelly, and S. Wagner, "Explaining math: Gesturing lightens the load," *Psychological Science*, 2001, 12(6): 516–522.
16. S. D. Kelly, A. Özyürek, and E. Maris, "Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension," *Psychological Science*, 2010, 21(2): 260–267.
17. K. G. Munhall, J. A. Jones, D. E. Callan, T. Kuratate, and E. Vatikiotis-Bateson, "Visual prosody and speech intelligibility: Head movement improves auditory speech perception," *Psychological Science*, 2004, 15(2): 133–137.
18. N. Ambady, F. J. Bernieri, and J. A. Richeson, "Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream," *Advances in Experimental Social Psychology*, 2000, 32: 201–271.
19. P. Borkenau, N. Mauer, R. Riemann, F. M. Spinath, and A. Angleitner, "Thin slices of behavior as cues of personality and intelligence," *Journal of Personality and Social Psychology*, 2004, 86(4): 599–614.
20. L. A. Zebrowitz and J. M. Montepare, "Social psychological face perception: Why appearance matters," *Social and Personality Psychology Compass*, 2008, 2(3): 1497–1517.
21. T. Bänziger, M. Mortillaro, and K. R. Scherer, "Introducing the Geneva multimodal expression corpus for experimental research on emotion perception," *Emotion*, 2012, 12(5): 1161–1179.
22. T. J. Clarke, M. F. Bradshaw, D. T. Field, S. E. Hampson, and D. Rose, "The perception of emotion from body movement in point-light displays of interpersonal dialogue," *Perception*, 2005, 34(10): 1171–1180.
23. N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture," *Emotion*, 2012, 12(5): 1085–1101.
24. F. Pollick, H. Paterson, A. Bruderlin, and A. Sanford, "Perceiving affect from arm movement," *Cognition*, 2001, 82: B51–B61.
25. K. R. Scherer, "Expression of emotion in voice and music," *Journal of Voice*, 1995, 9(3): 235–248.
26. J. C. Thoresen, Q. C. Vuong, and A. P. Atkinson, "First impressions: Gait cues drive reliable trait judgements," *Cognition*, 124(3): 261–271.
27. M. Koppensteiner, P. Stephan, and J. P. M. Jäschke, "More than words: Judgments of politicians and the role of different communication channels," *Journal of Research in Personality*, 2015, 58: 21–30.
28. R. S. Kramer, I. Arend, and R. Ward, "Perceived health from biological motion predicts voting behavior," *Quarterly Journal of Experimental Psychology*, 2010, 63(4): 625–632.
29. C. A. Klostad, R. C. Anderson, and S. Peters, "Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women," *Proceedings of the Royal Society of London B: Biological Sciences*, 2012, 297(1738): 2698–2704.
30. C. C. Tigue, D. J. Borak, J. J. O'Connor, C. Schandl, and D. R. Feinberg, "Voice pitch influences voting behavior," *Evolution and Human Behavior*, 2012, 33(3): 210–216.
31. M. Mehu and L. van der Maaten, "Multimodal integration of dynamic audio-visual cues in the communication of agreement and disagreement," *Journal of Nonverbal Behavior*, 2014, 38(4): 569–597.
32. D. V. Shah, A. Hanna, and E. P. Bucy *et al.*, "Dual screening during presidential debates: Political nonverbals and the volume and valence of online expression," *American Behavioral Scientist*, 2016, 60(14): 1816–1843.
33. D. V. Shah, A. Hanna, E. P. Bucy, C. Wells, and V. Quevedo, "The power of television images in a social media age: Linking biobehavioral and computational approaches via the second screen," *Annals of the American Academy of Political and Social Science*, 2015, 659(1): 225–245.
34. E. P. Bucy, "Emotional and evaluative consequences of inappropriate leader displays," *Communication Research*, 2000, 27(2): 194–226.
35. L. Bergen, T. Grimes, and D. Potter, "How attention partitions itself during simultaneous message presentations," *Human Communication Research*, 2005, 31(3): 311–336.

36. A. Lang, "Defining audio/video redundancy from a limited-capacity information processing perspective," *Communication Research*, 1995, 22(1): 86–115.
37. B. De Gelder and P. Bertelson, "Multisensory integration, perception and ecological validity," *Trends in Cognitive Sciences*, 2003, 7(10): 460–467.
38. J. Navarra, A. Alsius, S. Soto-Faraco, and C. Spence, "Assessing the role of attention in the audiovisual integration of speech," *Information Fusion*, 2010, 11(1): 4–11.
39. L. A. Ross, D. Saint-Amour, V. M. Leavitt, D. C. Javitt, and J. J. Foxe, "Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments," *Cerebral Cortex*, 2006, 17(5): 1147–1153.
40. R. A. Johnstone, "Multiple displays in animal communication: 'Backup signals' and 'multiple messages,'" *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 1996, 351(1337): 329–338.
41. C. Rowe, "Receiver psychology and the evolution of multicomponent signals," *Animal Behavior*, 1999, 58(5): 921–931.
42. K. Grammer, V. Filova, and M. Fieder, "The communication paradox and possible solutions," in *New Aspects of Human Ethology*, K. Atzwanger, K. Grammer, K. Schäfer, and A. Schmitt, eds. (New York: Springer, 1996), pp. 91–120.
43. M. Mehu, "The integration of emotional and symbolic components in multimodal communication," *Frontiers in Psychology*, 2015, 6: 7–11.
44. L. L. Carli, S. J. LaFleur, and C. C. Loeber, "Nonverbal behavior, gender, and influence," *Journal of Personality and Social Psychology*, 1995, 68(6): 1030–1041.
45. M. Koppensteiner, P. Stephan, and J. P. M. Jäschke, "Moving speeches: Dominance, trustworthiness and competence in body motion," *Personal and Individual Differences*, 2016, 94: 101–106.
46. M. Koppensteiner and K. Grammer, "Motion patterns in political speech and their influence on personality ratings," *Journal of Research in Personality*, 2010, 44(3): 374–379.
47. M. Koppensteiner, P. Stephan, and J. P. M. Jäschke, "Shaking takete and flowing maluma: Non-sense words are associated with motion patterns," *PLOS ONE*, 2016, 11(3): e0150610, doi:10.1371/journal.pone.0150610.
48. R. Ma, "Parametric speech emotion recognition using neural network," in *Proceedings of the 5th International Symposium on Neural Networks*, F. Sun, J. Zhang, J. Cao, and W. Yu, eds. (Berlin: Springer-Verlag, 2008).
49. Team RC, *R: A Language and Environment for Statistical Computing* (Vienna: R Foundation for Statistical Computing, 2013).
50. J. W. Johnson, "A heuristic method for estimating the relative weight of predictor variables in multiple regression," *Multivariate Behavioral Research*, 2000, 35(1): 1–19.
51. E. Krahmer and M. Swerts, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *Journal of Memory and Language*, 2007, 57(3): 396–414.
52. J. Van den Stock, R. Righart, and B. De Gelder, "Body expressions influence recognition of emotions in the face and voice," *Emotion*, 2007, 7(3): 487–494.
53. J. K. Burgoon, "Interpersonal expectations, expectancy violations, and emotional communication," *Journal of Language and Social Psychology*, 1993, 12(1/2): 30–48.
54. R. Awamleh and W. L. Gardner, "Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance," *Leadership Quarterly*, 1999, 10(3): 345–373.
55. Z. H. Gong and E. P. Bucy, "When style obscures substance: Visual attention to display appropriateness in the 2012 presidential debates," *Communication Monographs*, 2016, 83(3): 349–372.