Range- and domain-specific exaggeration of facial speech

Harold C. H. Hill Department of Vision Dynamics, Human Information Science Laboratories, ATR International, Kyoto, Japan

Nikolaus F. Troje Ruhr-Universität, Bochum, Germany, & Queen's University, Ontario, Canada

Alan Johnston Department of Psychology, University College London, London, United Kingdom

Is it possible to exaggerate the different ways in which people talk, just as we can caricature their faces? In this paper, we exaggerate animated facial movement to investigate how the emotional manner of speech is conveyed. Range-specific exaggerations selectively emphasized emotional manner whereas domain-specific exaggerations of differences in duration did not. Range-specific exaggeration relative to a time-locked average was more effective than absolute exaggeration of differences from the static, neutral face, despite smaller absolute differences in movement. Thus, exaggeration is most effective when the average used captures shared properties, allowing task-relevant differences to be selectively amplified. Playing the stimuli backwards showed that the effects of exaggeration were temporally reversible, although emotion-consistent ratings for stimuli played forwards were higher overall. Comparison with silent video showed that these stimuli also conveyed the intended emotional manner, that the relative rating of animations depends on the emotion, and that exaggerated animations were always rated at least as highly as video. Explanations in terms of key frame encoding and muscle-based models of facial movement are considered, as are possible methods for capturing timing-based cues.

Keywords: animation, biological motion, exaggeration, facial speech

Introduction

Biological motion, especially facial motion, provides a rich and subtle source of information for many tasks including recognition, facial speech, and the perception of emotion. The aim of this paper is to use methods for movement exaggeration to provide clues to how these different sources of information are encoded in movement. In particular, we focus on the differences associated with saying the same sentence with different emotional manners. We first introduce methods of exaggeration and then describe how they were applied to our particular task.

Effective facial caricatures, similar to those produced by hand, can be generated automatically by exaggerating the differences between individual faces and an average face (Brennan, 1985; Rhodes, Brennan, & Carey, 1987). These exaggerations can be better recognized than the originals (Rhodes et al., 1987), suggesting that faces are encoded as deviations from a stored prototype. Automatic exaggeration with respect to the dimensions of emotional expression (Calder, Young, Rowland, & Perrett, 1997), sex, attractiveness (Perrett et al., 1998; Perrett, May, & Yoshikawa, 1994), and age (Burt & Perrett, 1995) have also been successful. Some of these studies exaggerated visual texture as well as shape-based information (Rowland & Perrett, 1995; reviewed Rhodes, 1996). The underlying principal of exaggerating differences from the average has been effectively applied to motion data with differences parameterized spatially (Pollick, Fidopiastis, & Braden, 2001), temporally (Hill & Pollick, 2000), and spatiotemporally (Giese, Knappmeyer, Thornton, & Bülthoff, 2002). In this paper, we use both exaggeration of spatial range and exaggeration of variations in timing within the domain of a particular movement to provide clues to the encoding of facial movement.

Exaggeration is a fundamental principal of traditional animation and can be defined as "accentuating the essence of an idea via the design and the action" (Lasseter, 1987; Thomas & Johnston, 1981). Automatic facial caricature can be construed as exaggeration of design whereas motion exaggeration constitutes exaggeration of action. In both cases, the essential information is the information that differs from the average and that is accentuated by scaling.

Automatic exaggeration can be broken down into four stages: parameterization of the movement or shape, establishing correspondences, averaging the examples available, and exaggerating the differences between individual examples and the average.

The parameterization of shape is relatively straightforward, especially when markers are used, as it can simply be based on 2-D or 3-D positions. Many approaches to the

doi: 10.1167/5.10.4 Received September 26, 2003; published December 14, 2005 ISSN 1534-7362

parameterization of movement have been suggested in computer graphics (Amaya, Bruderlin, & Calvert, 1996; Bruderlin & Williams, 1995; Unuma, Anjyo, & Takeuchi, 1995) and computer vision (Essa & Pentland, 1997; Giese & Poggio, 2000; Troje, 2002a, 2002b; Yacoob & Black, 1999) for the purposes of synthesis and recognition. These approaches seek to represent movement linearly within vector spaces or, for periodic motions, Fourier-like harmonic spaces. Effective parameterization of both movement and shape are central to many problems including recognition and synthesis as well as exaggeration. A primary aim of this paper is to provide clues to how the human visual system parameterizes facial movement.

To establish correspondences, equivalent points have to be identified for the different exemplars. Again the use of markers simplifies this problem, at least to the extent that the markers can be placed in equivalent positions defined in terms of anatomical landmarks. Movement complicates the situation, as correspondences have to be temporal as well as spatial. For biological motion, this is complicated by the fact that physical time is nonlinearly related to the time scales of biological movements (Neri, Morrone, & Burr, 1998). Sequences are often of different lengths, and even after normalizing for total duration equivalent events may take place at different times with the results that averaging will tend to smooth out extreme positions. We follow Ramsay & Silverman (1997) in referring to this variation as domain-specific variation and use one of their listed methods, landmark alignment, in order to establish temporal correspondences. The variation in amplitude remaining after landmark alignment is referred to as range-specific variation.

In the experiments reported, we used motion capture data recorded from 10 people saying four sentences in four different ways: happily, sadly, angrily, and neutrally. This allowed us to limit the problem of establishing correspondences to different versions of the same sentence. The landmarks used were defined in terms of maxima in the position of the jaw marker, principally for the pragmatic reason that it proved relatively easy to identify equivalent maxima in these trajectories for different versions of the same sentence. These vertical maxima of the jaw are associated with consonant closures during speech (Vatikiotis-Bateson & Ostry, 1995). For example, the peaks in Figure 1 correspond to articulation of the consonants m (twice), f, and d in the sentence "I'm almost finished". The number of peaks is a function of the sentence being said and independent of manner. Principal components analysis of position data during speech shows that jaw movement constitutes the primary source of variation for facial movement during speech.

After identifying the timing of such events, it is possible to use resampling with interpolation to align the sequences. Both range- and domain-specific variation can then be averaged: range-specific variation by averaging relative position after the alignment of landmarks, and domain-specific variation by averaging the timings of those landmarks. After average range and domain values are established, differences from the average can be exaggerated. The effects of range- and domain-specific exaggeration are shown in Figure 2. Exaggeration of range corresponds to spatial exaggeration of movement relative to a time-locked average on a frame-by-frame basis (Pollick et al., 2001; Pollick, Hill, Calder, & Paterson, 2003), and exaggeration of domain-specific variations corresponds to temporal exaggeration of segment durations (Hill & Pollick, 2000).

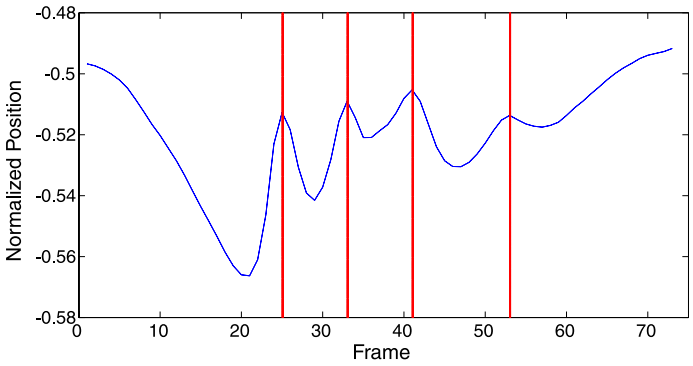


Figure 1. Normalized vertical jaw position for an example sentence. Vertical lines indicate the landmark peaks used, together with start and end points, for temporal alignment prior to averaging. The *x*-axis indicates time as frame number and the *y*-axis corresponds to vertical position normalized so that the difference between the highest and lowest marker in the neutral starting position was 1. This is an example of the sentence "I'm almost finished" said neutrally.

These techniques have been used previously to generate stimuli for perceptual experiments. For example, walking can be parameterized as eigen postures and the change in those postures over time modeled in terms of sine functions (Troje, 2002a). It is then possible to classify and exaggerate different styles of walking, for example, the differences between males and females, within the space defined. It was also possible to present structural and timing information independently and show that the latter was more informative than the former for categorization judgments by humans. The use of morphable models within a linear space has also been applied to face movements, and exaggeration of individual differences has been shown to enhance the recognition of identity (Giese et al., 2002).

In the experiments reported here, we focused on the differences associated with saying the same sentence in different ways. This results in variation on top of to the shared movement determined by the words spoken for each sentence. In everyday speech, such prosodic variation can change the manner and even the meaning of a sentence, without any change in the actual words

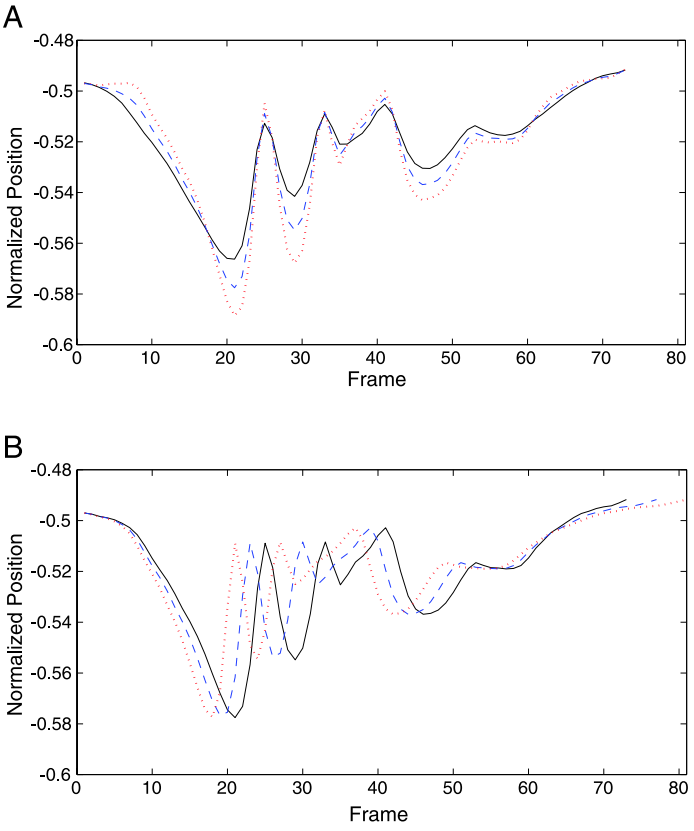


Figure 2. (A) Range-specific (spatial) exaggeration: Average vertical jaw position (black, solid) with average (blue, dashed) and exaggerated (red, dotted) angry versions of ''I'm almost finished''. Angry sequences have been temporally aligned with the grand average for illustrative purposes but were shown with their original timings in the experiment. (B) Domain-specific (temporal) exaggeration: The angry version of ''I'm almost finished'' with average timing (black, solid), its original timing (dashed, blue), and exaggerated timing (dotted red).

spoken. These variations have attracted considerable attention within the fields of auditory speech perception, production, and synthesis (Murray & Arnott, 1993), but until recently they have been largely neglected in the visual modality (but see Dohen, Loevenbruck, Cathiard, & Schwartz, 2004). Our primary interest was in whether and how differences in motion convey information in this context, and we restricted our stimuli to movement-based information alone. This was achieved by animating an average head with an average texture so that the only information available is derived from the motion capture data used to drive the animations (Hill & Johnston, 2001).

It is known that emotional expressions can be perceived visually from stimuli largely limited to movement information (Bassili, 1978, 1979) as well as static faces (Ekman, 1982). Similarly, movement alone can provide useful information for facial speech (Rosenblum, Johnson, & Saldaña, 1996; Rosenblum & Saldaña, 1996). The case of speech is particularly interesting because emotional manner is conveyed at the same time as speech content and person-specific information in both the auditory and visual channels. The close connection between faces and voices allows us to know whether someone is smiling or frowning from the sound of their voice alone, for example, on the telephone (Auberge & Cathiard, 2003; Tartter, 1994).

In this paper our primary concern was not the perception of emotion per se, but how variations in the basic motion determined by the words spoken convey manner. This is intended as a step in understanding an aspect of the very broad questions concerning the perception of biological, in particular facial, motion. To this end, observers were presented with silent animations and asked to categorize and rate the emotional intensity of the facial speech portrayed. Responses were measured as a function of the amount of range- and domain-specific exaggeration. Underlying this was the assumption that the effectiveness of exaggeration depends on the extent to which it enhances the perception and encoding of the information used for the task. The general methods will be described first, followed by the results of experiments comparing different types of exaggeration. Control experiments, including comparison with video stimuli, will then be reported before the results are integrated, conclusions drawn, and future directions considered.

General methods

Materials

The experiments reported here used marker-based motion capture recordings of facial movements made using an Oxford Metrics Vicon Motion Capture system with eight cameras at 120 Hz. Forty 2-mm diameter retro-reflective markers were placed on the face as shown in Figure 3. The three translations and three rotations for rigid motion of the head were estimated from the markers placed on the temples, the nose, and the top of the forehead. The remaining nonrigid movement of the face markers was then expressed relative to this rigid reference. Noise was present in the data due to spurious reflections. We reduced this noise by reconstructing the data from the first seven principal components derived from analysis of the vectors containing the coordinates for all the markers in each frame. While glitches remained, these were not manually edited in the interest of maintaining consistent treatment of all data.

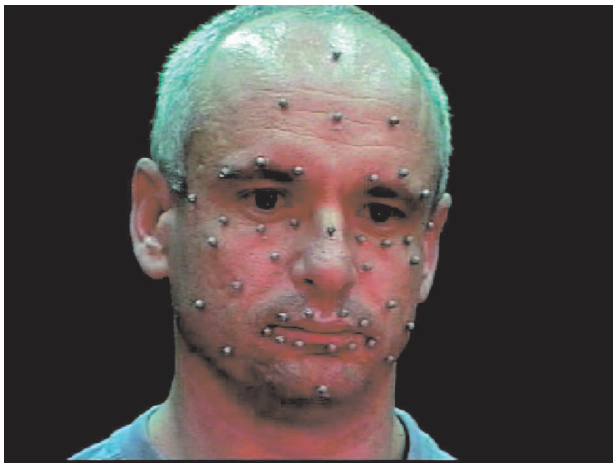


Figure 3. Marker placement for motion capture.

The recordings were of 10 fluent English speakers recruited from around the Ruhr University, Bochum. All said four sentences in four different ways: normally, happily, sadly, and angrily. The sentences were as follows:

1. What are you doing here?
2. That's what I thought.
3. I'm almost finished.
4. Look at that picture.

For each of the four sentences, we used corresponding peaks in the vertical position of the jaw as landmarks to temporally align the 40 different examples (four versions by 10 speakers) of that sentence. This was achieved by resampling each trajectory between landmarks by means of cubic spline interpolation, which resulted in corresponding peaks in the vertical position of the jaw marker all occurring at the same point (Figure 1). This temporal alignment allows calculation of average positions for all markers in every frame and generation of a grand average movement for each of the sentences used. Numerically manipulated position data were used to animate an average 3-D head model (Vetter & Troje, 1997) using methods described previously (Hill & Johnston, 2001). An example animation of the grand average for a sentence is shown in Movie 1.

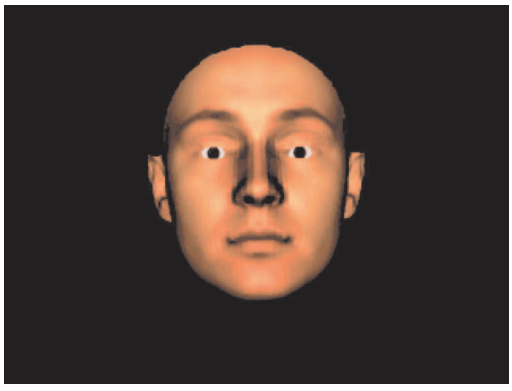
We also calculated the mean happy, sad, angry, and neutral averages separately, and it was exaggerations of these that were used as stimuli. Calculation of both grand and emotional averages was done in terms of position relative to the neutral starting frame rather than absolute position. This was to ensure that it was differences in movement rather than differences in marker placement or underlying shape that were exaggerated.

For range-specific exaggerations, relative position in each frame of the time-aligned sequence was first averaged and exaggerated. Subsequently, marker coordinates were again resampled so as to return landmark peaks to the original timings for that emotion. An example of a spatial exaggeration is shown in Movie 2.

Domain-specific exaggerations were produced by exaggerating differences in the timing of landmark peaks for emotional sequences relative to the grand average timings. Exaggerated sequences were generated by interpolation to give the required number of frames between peaks. Examples of domain-specific exaggerations are shown in Movie 3, and Figure 2 illustrates the effects of range- and domain-specific exaggeration on the vertical position profile of the jaw marker used for segmentation.

Observers

Observers for Experiments 1a and b and the video control experiment were recruited from foreign and Japanese staff at ATR, and those for Experiment 2 and the backwards control from the Ruhr University, Bochum.



Movie 1. The grand average for "I'm almost finished".

Movie 2. Examples of +1/2, +1, +2, and +3 range-specific exaggerations of "look at that picture" said happily. +1 corresponds to the average for that emotion.

Movie 3. Examples of +1/2, +1, and +2 domain-specific exaggerations of "look at that picture" said sadly.

Design

Within-subjects factorial designs were used for all experiments. The dependent variable was always 9-point ratings of happiness, sadness and anger, with rating scale included as a within-subjects factor. Although a far from perfect measure, ratings indicate the observer's categorization while also providing a more sensitive measure of differences between conditions. They have been widely used in similar experiments (e.g., Calder et al., 1997; Pollick et al., 2003). Intended emotion was also a within-subjects factor, and this was expected to interact with the rating scale factor—ratings of happy stimuli would be expected to be rated higher for happiness than for other emotions and so forth. Other factors involved the level and type of exaggeration and details are given for the separate experiments.

Procedure

Observers were instructed that they would be shown computer animations and asked to rate how happy, sad and angry each looked on a nine-point scale, with 9 indicating very and 1 not at all. They were told that some of the examples would be neutral, and that they should indicate these cases by giving low ratings on all of the scales. Observers could view each animation as many times as they wished, before responding in a text box using the number keys. In each case, entering the final rating initiated presentation of the subsequent trial. The order of stimuli was fully randomized for each observer.

Experiment 1a: Range-specific exaggeration

In this experiment observers rated the emotional intensity of range-specific exaggerations.

Design

For this experiment, a series of 64 animations were rendered. They comprised all possible combinations of the four different sentences, four emotions (neutral, happy, sad, and angry), and four different levels of exaggeration. The levels of exaggeration used were 1/2, 1, 2 and 3 times the difference between the grand average position for each frame and the average for the emotion. The level 1/2 corresponds to an anti-caricature falling half way between the grand average sequences and the average for an emotion, while 1 corresponds to the unexaggerated emotion and 2 and 3 to exaggerations of the emotion. After collapsing across sentence, this gave a 3 (Rating scale) × 4 (Emotion) × 4 (Exaggeration level) within-subjects design. The 64 animation sequences for each observer were presented in a randomized order.

Results and discussion

The effects of range-specific exaggeration on ratings of happiness, sadness, and anger are shown in Figure 4. Exaggeration increased intensity ratings for the scale corresponding to the emotion exaggerated; that is, happy ratings of happy sentences, sad ratings of sad sentences, and angry ratings of angry sentences. There was little effect of exaggeration on the ratings of neutral stimuli apart from a slight increase in anger ratings at the highest level of exaggeration. This shows that observers were able to perceive the intended emotion from the stimuli and that this perception was enhanced by range-specific exaggeration.

This pattern of results was confirmed by a 3 (Scale) × 4 (Emotion) × 4 (Exaggeration level) within-subjects analysis of variance (ANOVA) as summarized in Table 1.

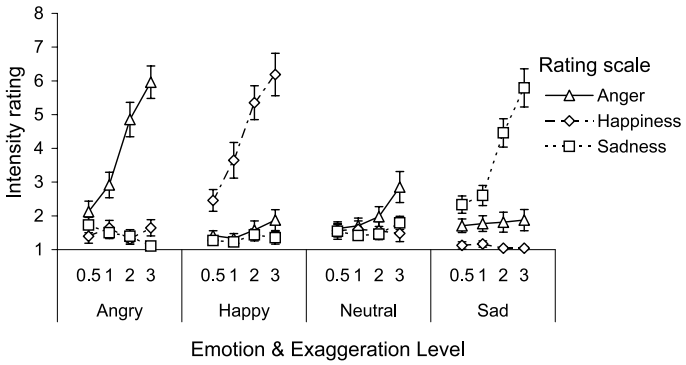


Figure 4. Emotional intensity ratings as function of range-specific exaggeration.

RS × EM × EL	RS × EL (Simple interaction)	EL (Simple, simple main effect)
$F(18,198) = 25.4, p \ll .05$	Angry stimuli: $F(6,66) = 29.3, p \ll .05$	Anger ratings: $F(3,33) = 43.9, p \ll .05$; *Sadness ratings*: $F(3,33) = 3.6, p < .05$
	Happy stimuli: $F(6,66) = 19.7, p \ll .05$	Happiness ratings: $F(3,33) = 52.2, p \ll .05$
	Sad stimuli: $F(6,66) = 26.6, p \ll .05$	Sadness ratings: $F(3,33) = 36.9, p \ll .05$
	Neutral stimuli: $F(6,66) = 2.0, p = .08$	Anger ratings: $F(3,33) = 4.5, p < .05$

Table 1. Summary of ANOVA for range-specific exaggeration. RS—rating scale, EM—emotion, EL—exaggeration level. Reductions in ratings with exaggeration are indicated with italics. Simple effects analysis was used to test all lower order interactions—these test whether main effects are significant at the different levels of the interacting variable(s) using a pooled error term (Nichols, 1993; Winer, Brown, & Michels, 1991). An alpha of .05 was used throughout with ≪ used when p values were an order or more below this. All effects and interactions not listed were $p > .1$.

In summary, range-specific exaggeration was effective in selectively increasing perceived emotional intensity for the intended emotion. In the case of neutral utterances, exaggeration did not make them appear emotional, except for a slight increase in rated anger at extreme levels of exaggeration. The negative effect of exaggeration on angry ratings of sad stimuli shows that exaggeration can also enhance discrimination by reducing false positives.

Experiment 1b: Domain-specific exaggeration

Using an analogous design to Experiment 1a, this experiment tested the effects of domain-specific exaggeration on rated emotional intensity.

Design

Details of the design were as for Experiment 1a, except that exaggerations were in terms of the durations of the intervals between temporal landmarks instead of spatial positions.

Results and discussion

Using domain- rather than range-specific exaggeration produced a quite different pattern of results, as can be seen in Figure 5. Compared to the effects of range-specific exaggeration, the effects of domain exaggeration were relatively small and not necessarily on the intended emotion.

The results of a three-way within-subjects ANOVA are summarized in Table 2.

Domain-specific exaggeration made happy and neutral stimuli look angrier, although it did not affect the perceived anger of angry stimuli. The manipulation had the effect of further reducing the durations of these types of sentence (see Table 3), and the resulting impression of increased "energy" of movement may have contributed to the perception of anger, although anger itself is actually longer in average duration. The "speeding up" of neutral stimuli also made them look less sad.

Timing is known to affect the perception of emotional expressions (Kamachi et al., 2001) and there were differences in the overall duration of different versions of the sentences that observers might have been sensitive to (see Table 3). However, exaggerating differences in duration did not, in general, enhance or interfere with the perception of perceived emotion. This, when contrasted with the results of Experiment 1a, suggests that the critical information for this task is the range of movement rather than the timing, or at least not the properties of timing captured by the method used here. Further discussion of this issue is left for the general discussion below.

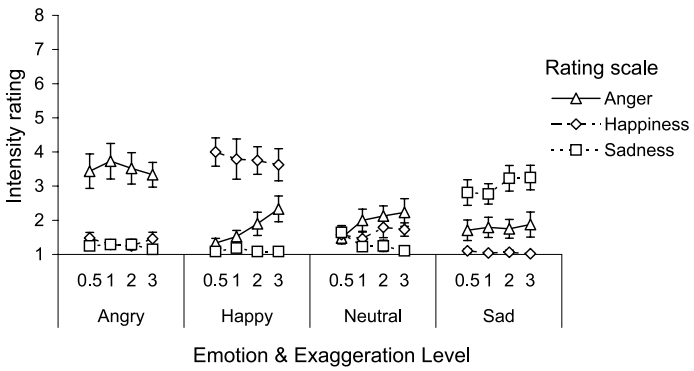


Figure 5. Emotional intensity ratings as function of domain-specific exaggeration.

RS × EM × EL	RS × EL (Simple interaction)	EL (Simple, simple main effect)
$F(18,198) =$ 1.9, $p < .05$	Angry stimuli, *ns*	
	Happy stimuli: $F(6,66) = 3.8$, $p \ll .05$	Anger ratings: $F(3,33) = 6.2$, $p \ll .05$
	Sad stimuli, *ns*	
	Neutral stimuli: $F(3,33) = 3.3$, $p < .05$	Anger ratings: $F(3,33) = 3.2$, $p < .05$; Sadness ratings: $F(6,66) = 2.8$, $p < .05$

Table 2. ANOVA summary for domain-specific exaggeration. For key see the caption to Table 1.

Experiment 2: Absolute and relative range exaggeration

Experiment 1a demonstrated the effectiveness of range-specific exaggeration relative to a time-locked average (hereafter relative exaggeration). However, for facial expressions, simply exaggerating differences from neutral has been found to be effective for both static photographs (Calder et al., 1997) and for dynamic point light displays (Pollick et al., 2003). In this experiment, we also exaggerated the absolute difference between each frame and the initial, static, and neutral position (hereafter absolute exaggeration). This is equivalent to a static spatial caricature of each frame (Brennan, 1985; Rhodes et al., 1987).

Absolute exaggeration is computationally far simpler than relative exaggeration as, for example, it does not require calculation of an average trajectory. If it were equally effective at exaggerating the salient information, this would suggest it reflects human encoding as well as favoring its use as an automatic method. A primary aim of this experiment was to compare the effectiveness of these two methods. Absolute exaggeration has the effect of exaggerating the absolute amplitude of movement, and so it provides a test of the extent to which this is the critical factor in exaggeration.

The difference between the two methods is best illustrated by sadness. This emotion is associated with reduced movement and thus relative exaggeration will tend to further reduce the amount of movement while absolute exaggeration will increase it. Examples of the two types of exaggeration for sadness can be seen in Movie 4.

	Sentence A	Sentence B	Sentence C	Sentence D
Angry	75	79	77	68
Happy	63	64	67	61
Neutral	67	65	62	63
Sad	82	72	88	74

Table 3. Timings in frames for the four sentences used according to manner.

Design

The design and procedure for this experiment was similar to that for Experiment 1 with the addition of type of exaggeration as a two-level factor. The number of levels of exaggeration was reduced from 4 to 3 (+1/2, +1, and +2). This resulted in a 2 (Type) × 3 (Rating scale) × 4 (Emotion) × 3 (Exaggeration level) within-subjects design. The four sentences in each condition gave a total of 96 trials, with responses to each of the three different rating scales on each.

Results and discussion

The results are plotted in Figures 6A and 6B. The 2 (Type) × 3 (Rating scale) × 4 (Emotion) × 3 (Exaggeration level) within-subjects ANOVA gave a four-way

Movie 4. Examples of +1/2 and +2 absolute and relative range exaggerations for “look at that picture” said sadly.

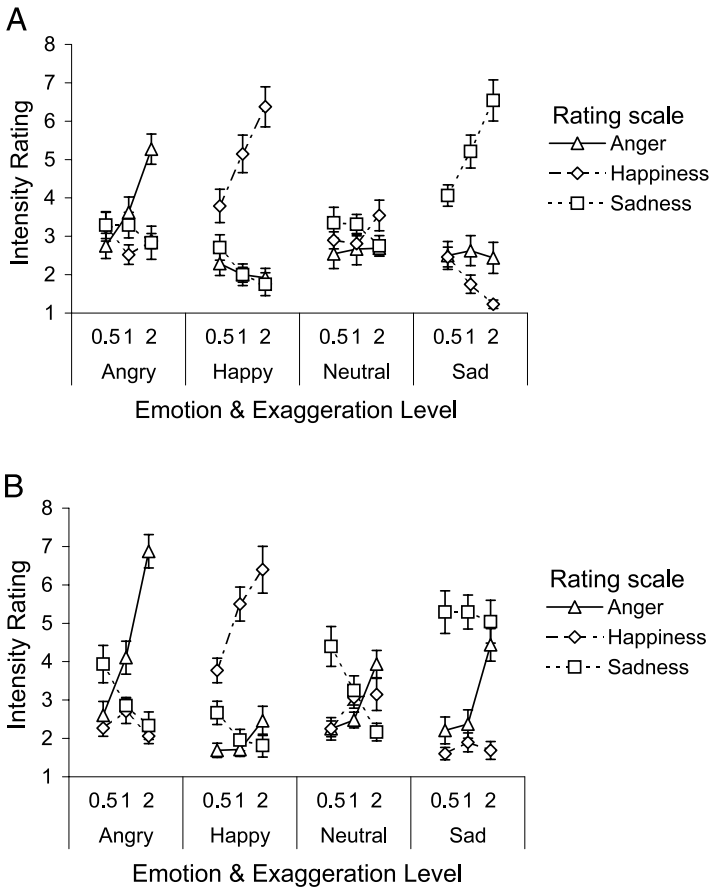


Figure 6. (A) Emotional intensity ratings for relative exaggeration. (B) Emotional intensity ratings for absolute exaggeration.

interaction, $F(12,132) = 2.5$, $p < .05$. This interaction shows that the pattern of results was different for the two types of exaggeration and so each was examined separately. Results are summarized in Table 4A for relative exaggeration and Table 4B for absolute exaggeration.

As can be seen in Figures 6A and 6B, the pattern of effects for absolute exaggeration was different, particularly for neutral and sad stimuli. This is confirmed by the significant four-way interaction summarized in Tables 4A and 4B.

The effects of relative exaggeration (summarized in Table 4A) were similar to those reported in Experiment 1a despite a different subject population and experimenter. One difference is that the observers in this experiment tended to give all stimuli a rating other than the default of one. This allowed negative effects of exaggeration on sadness rating of happy stimuli and happiness ratings of sad stimuli to become apparent. In this experiment, there was also no effect of exaggeration on angry (or any other) ratings of neutral stimuli, probably because the most extreme level of exaggeration was not used.

The effects of absolute exaggeration (summarized in Table 4B) were similar to relative exaggeration for happy and angry stimuli. For angry stimuli, the effects were more pronounced, probably because this manipulation produces a greater increase in movement (absolute exaggeration corresponds to exaggerating vertical distance relative to the starting point in Figure 2A rather than just the difference between black and blue trajectories). The pattern of results for neutral and sad stimuli was quite different, and not as intended. For the neutral stimuli, the increased movement associated with absolute exaggeration increased perceived anger and decreased perceived sadness. For the sad stimuli, exaggeration of absolute movement made the stimuli appear angrier.

To summarize, this experiment showed clear differences between relative and absolute exaggeration. The differences were most pronounced for neutral and sad stimuli. Increasing amplitude had the undesirable effect of increasing anger ratings for both neutral and sad stimuli, as well as desirably increasing ratings for angry stimuli. Decreasing amplitude increased sadness ratings for neutral and angry stimuli, leaving ratings for sad stimuli unaffected. These effects may, as with the effects of domain exaggeration reported in Experiment 1, reflect a general association of anger with increased movement and sadness with reduced movement. For angry and happy stimuli, the patterns of effects of the two types of exaggeration were similar, but exaggeration relative to a time-locked average might still be considered more efficient as smaller absolute differences in movement produced similar increases in perceived emotional intensity. The results suggest that the effect of exaggeration is not simply an effect of scaling the amplitude of movement, but rather of selectively scaling task-relevant differences.

The experiments above suggest that exaggeration of range-specific information relative to a time-locked average best enhances the critical information for this task. In the next experiment, we test the extent to which this information depends on the direction of movement. In the final experiment reported, we compare performance between animations and video.

Experiment 3a (control experiments): Stimuli played backwards

Playing stimuli backwards is an important control for motion experiments. It ensures that the same static frames are shown, and that kinematic properties are the same except for the change in sign. Despite leaving much information unchanged, this manipulation disrupts performance on some tasks, including recognition from motion (Hill & Johnston, 2001; Lander & Bruce, 2000).

	RS × EM × EL	RS × EL (Simple interaction)	EL (Simple, simple main effect)
A	$F(12,264) = 13.6$, $p \ll .05$	Angry stimuli: $F(4,44) = 8.7$, $p \ll .05$	Anger ratings: $F(2,22) = 23.0$, $p \ll .05$
		Happy stimuli: $F(4,44) = 17.6$, $p \ll .05$	Happiness ratings: $F(2,22) = 23.6$, $p \ll .05$; *Sadness ratings*: $F(2,22) = 5.6$, $p < .05$
		Sad stimuli: $F(4,44) = 17.9$, $p \ll .05$	Sadness ratings: $F(2,22) = 27.5$, $p \ll .05$; *Happiness ratings*: $F(2,22) = 12.8$, $p \ll .05$
		Neutral stimuli, *ns*	
B	$F(12,132) = 8.4$, $p \ll .05$	Angry stimuli: $F(4,44) = 26.8$, $p \ll .05$	Anger ratings: $F(2,22) = 60.0$, $p \ll .05$; *Sadness ratings*: $F(2,22) = 3.7$, $p < .05$
		Happy stimuli: $F(4,44) = 9.6$, $p \ll 0.05$	Happiness ratings: $F(2,22) = 21.2$, $p \ll .05$; *Sadness ratings:* $F(2,22) = 3.7$, $p < .05$
		Sad stimuli: $F(4,44) = 6.4$, $p \ll .05$	Anger ratings: $F(2,22) = 18.2$, $p \ll .05$
		Neutral stimuli: $F(4,44) = 12.4$, $p \ll 0.05$	Anger ratings: $F(2,22) = 21.2$, $p \ll .05$; Happiness ratings: $F(2,22) = 10.4$, $p \ll .05$; *Sadness ratings*: $F(2,22) = 10.3$, $p \ll .05$

Table 4. (A) ANOVA summary for relative range exaggeration found in Experiment 2. (B) ANOVA summary for absolute range exaggeration in Experiment 2. For key see the caption to Table 1.

Design

The design was the same as for Experiment 1 except that direction of movement, forwards or backwards, was included as an additional within-subjects factor. This resulted in a 2 (Direction) × 4 (Emotion) × 3 (Rating scale) × 4 (Exaggeration level) within-subjects design with a total of 128 trials.

Results

ANOVA gave a significant Emotion × Rating scale × Exaggeration interaction, $F(18,162) = 11.8$, $p \ll .05$. The pattern of results is shown in Figures 7A and 7B, was similar to Experiment 1 (see Figure 4), and did not depend on direction of movement, $p > .1$.

There were some significant third order interactions involving direction and the other factors. To unpack these effects of direction, the data for appropriate responses (that is, happy ratings of happy sentences, sad ratings of sad sentences, and angry ratings of angry sentences) were analyzed separately. This 2 Direction (forwards, backwards) × 3 Emotion (happy, sad or angry) × 4 Exaggeration (0.5,1,2,3) analysis resulted in independent main effects of Direction, $F(1,9) = 7.7$, $p < .05$; Emotion, $F(2,18) = 14.5$, $p \ll .05$; and Exaggeration, $F(3,27) = 47.1$, $p \ll .05$. The strong effect of exaggeration was as expected, appropriate ratings increasing with exaggeration as can be seen from Figures 7A and 7B. Overall appropriate ratings for stimuli were higher for stimuli played forwards, mean 4.8 (±0.3 *SEM*) than for stimuli played backwards, 4.4 (±0.3). There were also differences in the magnitudes of the appropriate ratings for the different emotions: happy 5.4 (±0.3), sad 4.8 (±0.4), and angry 3.8 (±0.4).

In summary, the effect of range exaggeration does not depend on the direction of movement, consistent with the cue exaggerated being temporally reversible. One possibility is that manner is encoded as a spatial modulation of the underlying motion signal. However, appropriate ratings were significantly higher overall for stimuli played forwards, showing that the pattern of change over time does affect perceived emotional intensity.

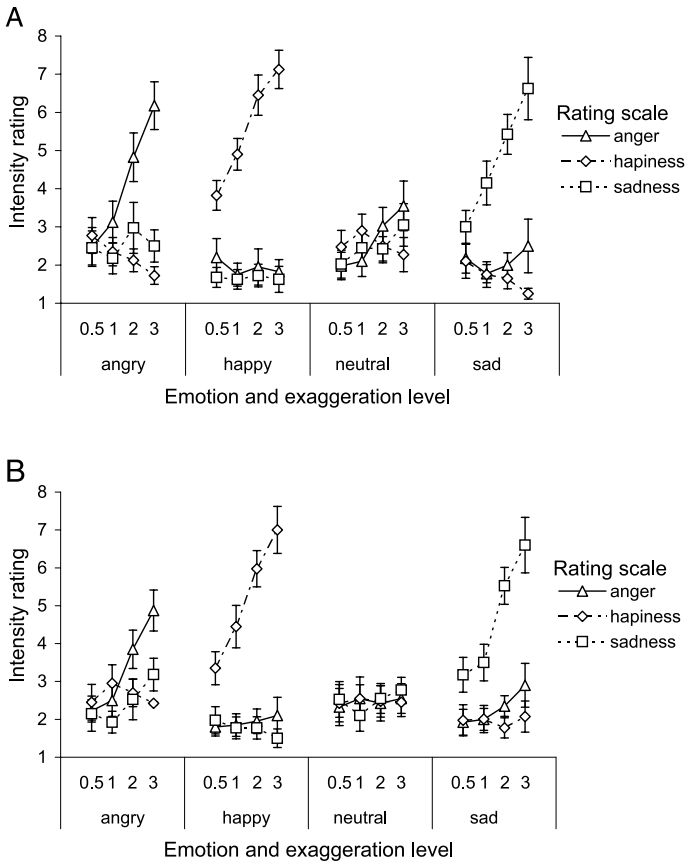


Figure 7. (A) Emotional intensity ratings as a function of exaggeration for stimuli played forwards. (B) Emotional intensity ratings as a function of exaggeration for stimuli played backwards.

Experiment 3b: Video control

The purpose of the final control experiment was twofold: firstly to test whether video of the original sequences conveyed the intended emotions, and secondly to test the perceived emotional intensity of exaggerated and unexaggerated animations relative to video. The design included unexaggerated, +1, and the most exaggerated, +3, range-specific animations from Experiment 1a, as well as video taken at the same time as the motion capture on which the animations were based. Animations and video clearly differ in many ways—in particular video contains much of the shape-based information that was deliberately excluded from the animations—but the pattern of ratings should still reveal how the perception of animations relates to more natural stimuli.

Materials

Animations were as described in the general methods. Videos were captured at the time of motion capture with a Sony mini-DV camera. Four versions (angry, happy, neutral, and sad) of each of four sentences were available for five males and five females. These were edited and encoded as 352 × 288 AVI movies with Mpeg-4 V2 encoding using Adobe Premiere. Frame rate was changed from 25 to 30 fps for compatibility with the animations. An example of the appearance of the stimuli can be seen in Figure 3. As can be seen, the motion capture markers were visible, as was red light from the motion capture cameras.

Design

The type of stimuli (video, and +1 or +3 range exaggerated animations) was included as a within-subjects factor. Exaggeration could not be included as a factor in the main analysis as it was not possible to show exaggerated videos. All 32 animations (4 sentences × 4 emotions × 2 levels of exaggeration) were shown to each observer together with 40 video sequences (one example of each emotion for each of the 10 people chosen at random from the four sentences available). Thus, this was a 3 (Type) × 4 (Emotion) × 3 (Rating scale) within-subjects design. Other details were as before.

Results and discussion

The results are plotted in Figure 8. As can be seen, video stimuli were perceived as having the emotion intended, although neutral stimuli tended to be seen as slightly sad or angry. This shows that the models spoke with the required manner and that observers can speech-read manner from silent video as well as animations. In this experiment, angry animations were not clearly recognized as such unless exaggerated. Exaggerated stimuli were always perceived as having greater emotional intensity than unexaggerated animations, with the exception of neutral stimuli.

As summarized in Table 5, ANOVA gave a three-way Type × Emotion × Rating scale interaction. The simple Emotion × Rating scale interactions were significant for all levels of Type. There were main effects of scale, consistent with the correct emotional categorization of the

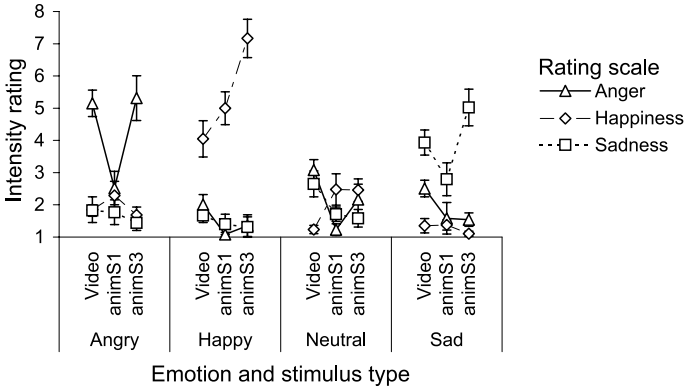


Figure 8. Emotional intensity ratings as a function of stimulus type and emotion.

Type × RS × EM	RS × EM (Simple interaction)	RS (Simple, simple main effect)
$F(12,132) = 11.1$, $p \ll .05$	Video stimuli: $F(6,66) = 32.9$, $p \ll .05$	Angry stimuli: $F(2,22) = 32.2$, $p \ll .05$; Happy stimuli: $F(2,22) = 12.1$, $p \ll .05$; Neutral stimuli: $F(2,22) = 17.8$, $p \ll .05$; Sad stimuli: $F(2,22) = 31.3$, $p \ll .05$
	Unexaggerated animations: $F(6,66) = 17.7$, $p \ll .05$	Angry stimuli, *ns*; Happy stimuli: $F(2,22) = 42.1$, $p \ll .05$; Neutral stimuli: $F(2,22) = 5.2$, $p < .05$; Sad stimuli: $F(2,22) = 4.2$, $p < .05$
	Exaggerated animations: $F(6,66) = 46.1$, $p \ll .05$	Angry stimuli: $F(2,22) = 29.0$, $p \ll .05$; Happy stimuli: $F(2,22) = 50.1$, $p \ll .05$; Neutral stimuli, *ns*; Sad stimuli: $F(2,22) = 41.8$, $p \ll .05$

Table 5. ANOVA summary for video control experiment. For key see Table 1 caption.

stimuli, for all emotions and all types of stimuli with the single exception of unexaggerated angry stimuli. Ratings of neutral stimuli also differed significantly, with the exception of exaggerated animations. These statistics are summarized in Table 5.

Animation appears to work least well (as compared to video) for angry stimuli, perhaps because it fails to capture some of the cues visible in the videos, for example, the narrowing of the eyes. However, exaggerating range does still enhance the cues to anger that are available. In general, video provides the shape-based information that was deliberately excluded from the animations, in addition to the movement-based cues that were of primary interest here. Video also does not suffer from problems associated with averaging, which will tend to result in smoothing, and thus reduce the perceived intensity of the unexaggerated animations.

In summary, this video control experiment demonstrated that the videos showed the intended emotions and that exaggerated animations were perceived as at least as intense as the video.

General discussion

The results show that it is possible to recover manner of speech from silent, motion-based animations as well as from video. Exaggerating the range of movement relative to a time-locked average appears to be an effective way of enhancing the information used for this task. On the other hand, exaggerating differences in durations had very different, and largely unintended, effects.

The results provide further evidence that the central principal of automatic caricature, exaggeration of differences from the average, can be applied to movement. However, it is also clear that this will only work if averages and exaggerations are calculated within an appropriate "space", as determined by the parameterization of movement used.

The effectiveness of exaggeration seems to be best explained in terms of its effects on the stimulus—it exaggerates diagnostic information while leaving common information, as captured by the average, unchanged. Thus, exaggeration of the spatial range of movement relative to a time-locked average works by leaving the shared properties of the movements, as determined by what is being said, unchanged while exaggerating the task-relevant differences associated with particular ways of saying the sentence. Simply exaggerating all movement (including shared patterns of movement) relative to the starting position is not as effective, as was shown by Experiment 2. We would not wish to suggest that an average is stored for every utterance. Instead merely that the time-locked average works by partialing out the motion information that is a function of what is being said thereby allowing the manner in which it is said to be selectively exaggerated. In this context, signal for one

task may constitute noise for another suggesting, for example, that and the average sentence shown in Movie 1 might be the best stimuli if the task involved determining what is being said.

The effectiveness of range-specific exaggeration suggests differences in the range of movements that contain important diagnostic information for this task. Changing the range of movement also changes kinematic properties, but these would also be affected by manipulations that change the durations of movements, including domain-specific exaggeration. For example, halving duration or doubling range would have equivalent effects on average velocity. The very different patterns of ratings for domain and range-specific exaggerations reported suggest that common kinematic properties are not the critical factor here.

The relationship between the two types of exaggeration on kinematic properties is further complicated as, in real speech, the duration and range of movement do not vary independently. The use of average emotions, derived from different people, avoided possible confounding effects of identity but may have further complicated the relationship between range and duration as it involves averaging the kinematics from individuals with different underlying dynamic properties such as jaw mass. This is a well-known problem in animation where, for example, movements from one actor have to be applied to a character at a different scale (Hodgins & Pollard, 1997).

Speech production has been modeled as a periodic attractor within a mass-spring system (Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Ostry & Munhall, 1985; Summerfield, 1987; Vatikiotis-Bateson & Kelso, 1993). In this case, the peak velocity and amplitude of the system are linearly related by stiffness. This relationship will be preserved by range-specific exaggeration, as it changes peak velocity and displacement proportionally, but disrupted by duration-specific exaggeration, which changes peak velocity without affecting maximum displacement. This suggests an alternative approach to exaggeration where the aim would be to infer the dynamic properties of the system from the kinematics, and then averaging and exaggerating differences within a dynamic parameter space.

The range-specific exaggeration technique used is also consistent with a key frame-based encoding scheme, within which movements are encoded in terms of static key frames. In classical pose-to-pose animation, spatial exaggeration is often applied to key frames, with intermediate frames "in-betweened", a process similar to interpolation (Laybourne, 1998). The timing of the key frames is of course also critical. The effectiveness of these animations suggests that they do "tap into" the brain's encoding of motion. The key frames may represent a form of eigen postures (Troje, 2002a), with motion encoded in terms of a trajectory within the space defined by the postures. Key frames, in the guise of snapshot neurons, are also a central component of a recent model of biological motion recognition (Giese & Poggio, 2003). With such a scheme, the effectiveness of range-specific exaggeration is interpretable as enhancing recognition of the form-based component of movement encoding. In the case of the animations used here, this form information would of necessity be of the type referred to as "motion-mediated structural information" (Troje, 2002a) or "motion-induced spatial information" (Knappmeyer, Thornton, & Bülthoff, 2003), which is the spatial differences resulting from applying different movements to the same initial starting shape.

Alternatively, range-specific exaggeration may be amplifying an underlying, fairly constant, spatial modulation associated with different manners of speech. The muscle movements associated with expression may be being modulating those associated with speech throughout the utterance. Happiness is associated with the raising of corners of the mouth and eyes, and sadness with their lowering, whereas anger is most apparent in the brow (Ekman, 1982). These expressions could introduce an essentially spatial modulation of the underlying motion signal associated the speech throughout the utterance. Relative exaggeration of spatial range would then be explained as enhancing the visibility of these modulations by exaggerating their extents. A muscle-based animation system (Lucero & Munhall, 1999; Parke & Waters, 1996; Waters, 1987; Waters & Terzopoulos, 1992), with exaggeration within a space defined in terms of muscle parameters, might best capture such effects.

The finding in Experiment 3a that the effect of exaggeration is temporally reversible is consistent with the above form-based accounts or with direction-independent kinematic or dynamic accounts. However, playing stimuli backwards would have been expected to disrupt the temporal patterning of the type we were seeking to capture with domain-specific exaggeration.

The unintended effects of domain-specific exaggeration reported in Experiment 2 contrast with previous work, where this type of exaggeration was found to enhance the recognition of identity from arm movements when range information was kept constant (Hill & Pollick, 2000). The current result is consistent with a previous study that did not show an advantage of exaggerating differences in the onset durations of facial expressions shown as point light displays (Pollick et al., 2003). Work on exaggerating individual differences in facial movement also reports better results if the durations of movement segments are not morphed but set to the average duration (Giese et al., 2002). Clearly one possible explanation of these contrasting results is that there is a fundamental difference between the encoding of arm and face movements, consistent with neuropsychological reports of such a dissociation (Campbell, Zihl, Massaro, Munhall, & Cohen, 1997). However, both arm and other body movements can, like face movements, be expressive of emotion

(Amaya et al., 1996; Pollick, Paterson, Bruederlin, & Sanford, 2001), and similar rules for the interpretation of movement might be expected to apply.

Absolute differences in timing have also been shown to affect the perception of facial expressions of emotion (Kamachi et al., 2001), and we would certainly not wish to deny the importance of timing in the perception of facial movement. However, a locally linear manipulation of duration does not appear sufficient to capture this important source of variation and more sophisticated methods are clearly needed. The landmark alignment method used may also have blurred timing cues. The landmark-defined segments also did not distinguish between movements associated with consonants and vowels, although different phonemes show different degrees of temporal compressibility.

In this work, there was no detrimental effect of exaggerating duration, although our domain-specific exaggerations are likely to have violated production constraints. We might have found such decrements if the task had involved recovering information about speech content, as this task depends more on determining the sounds being made than does the task of recovering manner.

An effective method for domain-specific exaggeration must preserve those properties that cannot vary, for example, transitions critical to the physics of speech production, while allowing those properties that can vary, for example, sustains and pauses, to vary. In theory, spatiotemporal exaggeration based on a dense set of perfect correspondences would do just that: Invariant properties would be the same for exemplars and the average and thus not affected by exaggeration while difference that did exist would be selectively enhanced. However, the uniqueness of each utterance makes such a perfect set of correspondence difficult to achieve, especially when features are present in one utterance that are absent from the other. For the human system, there is also the problem that movements must be perceived in real-time, without the luxury of having the complete trajectory available for analysis.

It will take time to fully determine the basis of the human parameterization of movement. The effectiveness of the simple range-specific exaggeration technique reported here provides clues, but not a final answer, to the form that this parameterization takes. Clearly the spatial extent of movement is critical, and it is not sufficient simply to exaggerate differences in duration while keeping range constant. Key frame, dynamic mass spring, and muscle-based models all appear to provide plausible explanations of the effectiveness of time-locked, range-specific exaggerations. As well as pursuing these methods, we are also looking at how people naturally exaggerate their speech, for example, when emphasizing a durational contrast, talking to a foreigner, speaking over noise, or caricaturing someone's mannerisms.

Acknowledgments

This research was supported the National Institute of Information and Communications Technology, the Engineering and Physical Science Research Council, and the Volkswagen Foundation. Eric Vatikiotis-Bateson provided many helpful comments and encouragement and Tobias Otto kindly ran the experiments in Bochum. Two anonymous reviewers provided many stimulating suggestions.

Commercial relationships: none.
Corresponding author: Harold Hill.
Email: hill@atr.jp.
Address: Human Information Science Laboratories, ATRi, Kyoto, 619-0288, Japan.

References

Amaya, K., Bruderlin, A., & Calvert, A. (1996). *Emotion from motion.* Paper presented at the Graphics Interface.

Auberge, V., & Cathiard, M. (2003). Can we hear the prosody of smile? *Speech Communication, 40,* 87–97.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology. Human Perception and Performance, 4*(3), 373–379. [PubMed]

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37*(11), 2049–2058. [PubMed]

Brennan, S. E. (1985). The caricature generator. *Leonardo, 18,* 170–178.

Bruderlin, A., & Williams, L. (1995). *Motion signal processing.* Paper presented at the SIGGRAPH: International Conference on Computer Graphics and Interactive Techniques.

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian males faces: Computer graphics manipulation of shape and colour information. *Proceedings of the Royal Society of London. Series B, 259,* 137–143. [PubMed]

Calder, A. J., Young, A. W., Rowland, D., & Perrett, D. I. (1997). Computer enhanced emotion in facial expression. *Proceedings of the Royal Society of London. Series B, B264,* 919–925. [PubMed]

Campbell, R., Zihl, J., Massaro, D., Munhall, K., & Cohen, M. M. (1997). Speechreading in the akinetopsic patient. L.M. *Brain, 120*(Pt 10), 1793–1803. [PubMed]

Dohen, M., Loevenbruck, H., Cathiard, M.-A., & Schwartz, J.-L. (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication, 44*, 155–172.

Ekman, P. (1982). *Emotion in the human face.* Cambridge: Cambridge University Press.

Essa, I. E., & Pentland, A. P. (1997). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 757–763.

Giese, M., Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2002). Spatiotemporal exaggeration of complex biological movements [Abstract]. *Perception, 31*(Supplement), 61.

Giese, M., & Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision, 38*(1), 59–73.

Giese, M., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192. [PubMed]

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11*(11), 880–885. [PubMed]

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science, 11,* 223–228. [PubMed]

Hodgins, J. K., & Pollard, N. S. (1997). *Adapting simulated behaviours for new characters.* Paper presented at the Sigraph '97, Los Angeles.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expression. *Perception, 30,* 875–887. [PubMed]

Kelso, J. A., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustic Society of America, 77,* 266–280. [PubMed]

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43,* 1921–1936. [PubMed]

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12*(4), 259–272.

Lasseter, J. (1987). Principal of traditional animation applied to 3D computer animation. *Computer Graphics, 21*(4), 35–44.

Laybourne, K. (1998). *The animation book.* New York: Three Rivers Press.

Lucero, J. C., & Munhall, K. (1999). A model of facial biomechanic for speech production. *Journal of the Acoustic Society of America, 106,* 2834–2843. [PubMed]

Murray, I., & Arnott, J. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustic Society of America, 2,* 1097–1108. [PubMed]

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395,* 894–896. [PubMed]

Nichols, D. P. (1993). Testing simple effects in MANOVA. Retrieved 23/6/2003 from ftp://ftp.spss.com/pub/spss/statistics/nichols/articles/simple.txt.

Ostry, D. J., & Munhall, K. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America, 77,* 640–648. [PubMed]

Parke, F. I., & Waters, K. (1996). *Computer facial animation.* Wellesley, Massachusetts: A K Peters Ltd.

Perrett, D. I., Less, K. J., Penton Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394,* 884–887. [PubMed]

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgements of female attractiveness. *Nature, 368,* 239–242. [PubMed]

Pollick, F. E., Fidopiastis, C. M., & Braden, V. (2001). Recognizing the style of spatially exaggerated tennis serves. *Perception, 30,* 323–338. [PubMed]

Pollick, F. E., Hill, H., Calder, A. J., & Paterson, H. (2003). Recognizing expressions from spatially and temporally modified movements. *Perception, 32,* 813–826. [PubMed]

Pollick, F. E., Paterson, H., Bruederlin, A., & Sanford, A. J. (2001). Perceiving effect from arm movement. *Cognition, 82,* B51–B61. [PubMed]

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis.* New York: Springer-Verlag.

Rhodes, G. (1996). *Superportaits.* Hove, East Sussex: Psychology Press.

Rhodes, G., Brennan, S. E., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473–497. [PubMed]

Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research, 39*(6), 1159–1170. [PubMed]

Rosenblum, R. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology. Human Perception and Performance, 22*(2), 318–331. [PubMed]

Rowland, D., & Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications, 15*(5), 70–76.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech processing. In R. Campbell (Ed.), *Hearing by eye: The psychology of lipreading* (pp. 3–51). Hove, UK: Erlbaum.

Tartter, V. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustic Society of America, 96*(4), 2101–2107. [PubMed]

Thomas, F., & Johnston, O. (1981). *Disney animation: The illusion of life* (1st ed.). New York: Abbeville Press.

Troje, N. F. (2002a). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*(5), 371–387, http://journalofvision.org/2/5/2/, doi:10.1167/2.5.2. [PubMed] [Article]

Troje, N. F. (2002b). The little difference: Fourier-based gender classification from biological motion. In M. Lappe (Ed.), *Dynamic perception* (pp. 115–120). Berlin: Aka Verlag.

Unuma, M., Anjyo, K., & Takeuchi, R. (1995). *Fourier principals for emotion-based human figure animation.* Paper presented at the SIGGRAPH: International conference on computer graphics and interactive techniques.

Vatikiotis-Bateson, E., & Kelso, J. A. S. (1993). Rhythm type and articulatory dynamics in English, French, and Japanese. *Journal of Phonetics, 21,* 231–265.

Vatikiotis-Bateson, E., & Ostry, D. J. (1995). An analysis of the dimensionality of jaw motion in speech. *Journal of Phonetics, 23,* 101–117.

Vetter, T., & Troje, N. F. (1997). Separation of texture and shape in images of faces for image coding and synthesis. *Journal of the Optical Society of America. A, Optics, Image Science and Vision, 14,* 2152–2161.

Waters, K. (1987). *A muscle model for animating three-dimensional facial expression.* Paper presented at the SIGGRAPH.

Waters, K., & Terzopoulos, D. (1992). The computer synthesis of expressive faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 335*(1273), 87–93. [PubMed]

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principals in experimental design* (3rd ed.). New York: McGraw-Hill.

Yacoob, Y., & Black, M. J. (1999). Parameterized modeling of recognition of activities. *Computer Vision and Image Understanding, 73*(2), 232–247.