"MEOW! Okay, I shouldn't have done that":
Factors Influencing Vocal Performance through Animoji

Susan C. Herring,[1] Ashley R. Dainas,[2] Ying Tang[3]
Indiana University Bloomington[123]
herring@indiana.edu,[1] ardainas@iu.edu,[2] yt11@iu.edu[3]

Abstract

This study investigates how and to what extent Animoji are vocally performed in private messages to recipients of varying levels of intimacy. Animoji users matched seven Animoji with seven relationships at different degrees of social distance and recorded a message for each pairing. The messages of 17 participants were narrowly transcribed, coded for performance features, and analyzed by Animoji and relationship type. Participants tended to modify their speech in different ways (e.g., via prosody, accent, lexis) with different Animoji, although the modifications were not as extreme as in previously studied public Animoji performances. Relationship type appears to be a better predictor of performativity than Animoji overall, but there are notable exceptions.

Introduction

Animoji on the Apple iPhone are larger versions of such familiar emoji as the dog, the unicorn, and the pile of poop. Both types of icon are used playfully, and they are exchanged mostly between intimates (Herring et al. 2020c; Konrad et al. 2020). At the same time, Animoji differ from emoji in important respects. Animoji track a user's facial movements in real time and can be "worn" as dynamic masks in video chat and in recorded video clips. Moreover, "wearing" an Animoji is typically accompanied by speaking; this verbal component is absent from emoji and other graphicons such as emoticons, stickers, and GIFs (cf. Herring and Dainas 2017). Although Animoji vocalizations have so far received little attention in the graphicon literature, they constitute an intriguing object of study in their own right. How do people speak through Animoji?

In a recent study, Herring et al. (2020a) analyzed Animoji clips posted publicly to YouTube and Twitter and observed that vocalizations in those clips deviated from the user's normal speaking voice according to what the Animoji represents. Certain Animoji, such as the robot, appeared to trigger associated vocal stereotypes especially often (e.g., a monotone delivery and sound effects such as 'beep beep'). While some instances of this were clearly intended as entertaining

performances, other vocal modifications seemed to occur almost as unconscious reflexes, and some reproduced racial and ethnic stereotypes. These observations led us to ask: What about Animoji vocalizations in more prototypical contexts of use: in private communication with friends and family? Do users still modify their speech, and if so, in what ways? Does the degree to which a vocalization is performed, rather than spoken in a normal voice, vary according to the Animoji and the relationship of the sender to the recipient?

To address these questions, we conducted a study in which we had Animoji users match a set of seven Animoji with seven relationships at different degrees of social distance, then send the chosen Animoji in a text message to each "relationship" (in reality, to the researchers' phone). The participants were told that they could say anything in the messages and were given no instructions as to how to speak. The vocalizations in each Animoji clip were subsequently transcribed using a narrow transcription system to capture variations in prosody and vocal quality, and the transcription features were analyzed in relation to two independent variables: the relationship of the recipient to the sender and the Animoji used. The results reveal that even in a laboratory setting, the Animoji influenced speech production overall in the direction of playful performance, although the manner of performance varied by Animoji. At the same time, relationship type (as a proxy for social distance) better accounted for many features of vocal performativity than Animoji did. Messages to intimate recipients were generally the most performed, with the exception of close friends, with whom many performance features were avoided. Performance features were used more than expected with distant relationships, presumably to mitigate social awkwardness.

Relevant Background

Verbal Performance

Verbal performance is identity expression through manipulation of language (Bauman 1975) in which performers consciously or unconsciously modify their word choice and the

prosodic features of their speech – stress, pitch, intonation, volume, rhythm, etc. In one type of verbal performance, *stylization*, performers adopt the voice of another and repurpose it to suit their own objectives. Such stylization may invoke "altera personae" drawn from familiar socio-cultural repertoires, including archetypes, stereotypes, and well-known media characters (Coupland 2001). In the process, performers convey cultural and ideological meanings that often extend beyond the propositional content of the spoken words (Bauman 1975). Analyzing such performances helps identify the communicative means that are used as performative shorthand, as well as "the degree of intensity with which the performance frame operates" (Bauman 1975, p. 297).

In the Animoji video clips posted publicly to social media that Herring et al. (2020a) analyzed, the animators of the clips stylized their speech and modified their vocal quality, prosody, and lexis to perform different personae. They produced sounds and accents characteristics of the Animoji or Memoji (customizable Animoji that represent human heads) through which they were speaking, such as barking and panting for the Dog, flat pitch and beeping sounds for the Robot, and ethnic and regional accents for Memoji (e.g., African-American Vernacular English [AAVE] features for a generic "black mom"), as well as specific speech styles to represent specific persons or characters (e.g., raised pitch to voice the persona of a female user as a little girl). These results support the Proteus Effect (Yee and Bailenson 2007), according to which the avatar one adopts shapes one's behavior. The study further revealed a tendency for Animoji performances to lead, consciously or unconsciously, to social stereotyping. Finally, gender differences were identified in the frequency and intensity of performance of the Animoji clips, with men posting more clips and exaggerating their vocalizations more than women. However, those clips were shared publicly for entertainment purposes; the speech that accompanies private Animoji use could be quite different.

Voice Messages

Animoji video clips sent as text messages share similarities with both telephone voicemail and audio/video messaging through smartphone apps. All are asynchronous, technology-mediated, "one-sided social actions" (Alvarez-Caccamo and Knoblauch 1992) that can be construed as solo verbal performances. As an older phenomenon, the language of telephone voicemails has been much studied by linguists. Voicemails tend to contain unplanned speech (Ochs 1979) marked by false starts, redundancies, hedges, and pause fillers used to buy more time to plan what to say next, as well as humor and laughter which can be used to mitigate the awkwardness of speaking to a non-present addressee. They typically start with an opening sequence, where the caller identifies themselves, provides a ritualized greeting such as 'hello' or 'hey,' and mentions the caller's

name. The core portion provides information about why they are calling, followed by a closing sequence in which the caller terminates the call (Alvarez-Caccamo and Knoblauch 1992). In closing, callers often begin with a word like "ok" or "anyway" and/or falling, low tones, followed by a pause (Gold 1991). Similar patterns were observed in the Animoji clips recorded by our study participants.

Animoji clips also resemble asynchronous audio and video messages shared in multimodal smartphone-based social media apps such as WhatsApp and WeChat. Adoption and use of voice messaging have seen a substantial increase recently. Hass et al. (2020) were among the first to address this trend by surveying and interviewing smartphone users about their motives for using voice messaging. Apart from convenience and situational needs, an important motivation the users identified was to convey paralinguistic features that otherwise would be lost in written or graphical text messages, including prosody, pitch, volume, intonation, and inflection. These paralinguistic features add a personal touch to the message and help convey emotions and express ideas. The same appears to be true of Animoji messages.

Graphicons and Social Distance

Graphicons such as emoji and stickers are highly expressive and personal means of communication. As phatic communication, emoji and stickers are sent to amuse the recipient (Konrad et al. 2020; Pazil 2018), as icebreakers, and as a tool for self-expression in group chat (Lee et al. 2016). Graphicon use is especially associated with intimate contexts and the familiarity, trust, and self-disclosure associated with intimacy (Prager 2008; Reis and Shaver 1988; Wang 2016). An experimental study showed that just increasing the number of emoticons in a conversation led to higher perceived levels of intimacy (Janssen et al. 2014). Thus graphicons help establish, maintain, and manage relationships at different social distances, where social distance is "the perceived or desired degree of remoteness between a member of one social group and the members of another, as evidenced in the level of intimacy tolerated between them."[1] However, Konrad et al. (2020) posit that when graphicons are new, their use is usually restricted to very close relationships until the graphicon achieves more widespread adoption. Herring et al. (2020c) supported this proposition with evidence from self-reported Animoji use.

In interviews with Midwestern university students, Herring et al. (2020c) found that Animoji were sent most often to intimates – close friends, significant others, siblings, and sometimes to parents and extended family members. Animoji were not considered appropriate to send to more socially distant relationships such as a new friend or a teaching assistant. When asked to match seven Animoji with seven relationship types, users were thoughtful about which Animoji were appropriate to send to different recipients, with

the Poop considered the most transgressive and appropriate to send only to a significant other, close friend, or sibling. The Dragon and Robot were considered polite or respectful when sent to an older relative or TA, but "weird" and funny when sent to (male) close friends or new male friends, while the Dog and Monkey were considered "normal" and appropriate to send to almost anyone (Herring et al., 2020c). The present study analyzes data from the same interviews; however, our previous study did not analyze the actual messages sent, the speech that accompanied the messages, or how the messages compare with publicly shared Animoji clips such as those analyzed in Herring et al. (2020a).

Data and Methods

Data

The goal of this study is to investigate the characteristics of the speech that accompanies Animoji video clips sent through iMessage on the Apple iPhone, and to determine how that speech varies according to which Animoji is sent and to which relationship (degree of social distance) in a non-public setting. We conducted in-depth semi-structured interviews with self-identified Animoji users between November 2019 and March 2020. Participants were recruited from the population of a Midwestern university town who had an Apple iPhone X, XR, or 11 (hereafter referred to as an iPhone) and had some prior experience with Animoji. A total of 33 participants were interviewed face-to-face in a laboratory setting (19 females, 13 males, 1 nonbinary; age range 18-29; 55% native English speakers).

As part of the study, participants completed a matching task and sent Animoji video messages to hypothetical addressees. Participants were given two decks of cards, one depicting seven diverse Animoji characters – Dog, Cat, Monkey, Rabbit, Dragon, Robot, and Poop – and one depicting seven common relationships at varying degrees of intimacy/distance – significant other (SO), close friend (CF), sibling (S), older relative or mentor (OR), new female friend (NFF), new male friend (NMF), and teaching assistant (TA).[2] The participants were asked to choose the Animoji that they would send to a person in each relationship category and explain the reasons for their choices.

After matching the Animoji and Relationship cards, the participants recorded a 15-30[3] second, asynchronous video clip for each Animoji-relationship pair on their phone and sent it to the researcher's phone, as if they were sending it to the relationship for whom it was intended. They were told to use their imaginations and say anything they wanted in

the Animoji clips, and no instructions or models were provided as to how to speak. To give the participants privacy, the researcher left the room while the clips were recorded. The Animoji messages were later transcribed by a trained research assistant using a narrow transcription system loosely based on the Jeffersonian system favored for conversation analysis (Jefferson 2004) to capture variations in prosody and vocal quality along with laughter and other non-speech vocalizations (see Appendix).[4]

Because analysis of vocalizations is very labor intensive, in this study we analyzed a gender-balanced subset of the transcribed data consisting of 118 messages produced by 17 participants (8 F, 8 M, 1 nonbinary; 18-26 years old; 65% native English speakers). The lead author checked the transcriptions and divided the messages into utterances based mostly on intonation contours, resulting in 562 utterances. Utterances averaged 7.29 words (range: 1-25), and participants averaged 4.8 utterances (range: 1-10) per message.

Methods

We operationalized performance in terms of four categories: disfluencies, prosody, special vocalizations, and Animoji references. All but disfluencies were derived directly or indirectly from Herring et al. (2020a). *Disfluencies* are common features of speech production that are usually unplanned (Ochs, 1979) but that can be used intentionally for effect. They comprise pauses, pause fillers (such as 'um,' 'uh,' 'er'), and questioning intonation, which is associated with uncertainty and tentativeness (cf. Linneman 2012).

Herring et al. (2020a) coded 'degree of performance' impressionistically on a scale from 0 to 3. In this study, we broke performance down into individual measurable verbal behaviors. *Prosody* comprises lengthening of a sound, rising pitch, falling pitch, and emphasis. *Special vocalizations* include marked vocal qualities (such as vocal fry, sing-song) and language styles (such as AAVE), as well as Animoji Sounds and Colloquial Pronunciation. Animoji Sounds includes sounds typically made by an Animoji character, as well as other sounds that were produced as part of a performance of it (e.g., "Hee-hee-hee-hee" in the following utterance: "Hee-hee-hee-hee {*monkey sounds* ooh-ooh-ooh. Hihihihi}"); further, we counted the number of words produced using all those sounds.[5] Colloquial Pronunciations often occur as a feature of accent stylization (e.g., "Hopefully you'se doin' good" uttered with a vocal quality reminiscent of Bugs Bunny when speaking through the Rabbit Animoji).

Finally, *Animoji References* include two phenomena coded by Herring et al. (2020a): Mentioning the Animoji by Name (e.g., "I'm a poop!") and Mentioning Characteristics

of or Associations with the Animoji (e.g., in sending the Robot to a TA: "Get it, cuz we program these days"). To these we added a new feature that emerged from our data, Metamention of the Animoji as an object with affordances (e.g., "[bobbing head up and down] Oh look my ears move too.").

The three authors coded all the above features in a random contiguous sample of 50 utterances, resulting in interrater agreement scores above 95% for each coded feature. Subsequently, the remaining utterances were evenly divided and coded by the authors separately. In the Results section, the frequency of each feature is analyzed by relationship and Animoji. As the size of the dataset does not permit statistical analyses, all results are presented as descriptive statistics.

Hypotheses

The Animoji users we interviewed (see Herring et al. 2020c), some of whose messages we analyze in the present study, associated each of the seven Animoji with a relationship in the card sort task. The main associations are summarized in Table 1 below. The strongest associations are indicated with a larger check mark.

	SO	CF	S	OR	NFF	NMF	TA
Dog	✓✓ F	✓	✓	✓	✓	✓✓	✓
Monkey	✓✓ F	✓	✓	✓	✓✓	✓	
Poop	✓✓ M	✓✓	✓✓				
Dragon		✓		✓		✓✓	
Rabbit			✓	✓	✓✓		
Cat				✓	✓✓		✓
Robot						✓	✓✓

F especially females; M especially males

Table 1. Animoji-Relationship Matches (Herring et al. 2020c)

The study participants also identified the Animoji that they would be likely to send to those relationships in actuality, from most and least likely:

Dog/Monkey > Cat > Rabbit > Poop > Dragon > Robot

We follow this order in the rest of the tables in this paper.

In addition, the participants ranked the relationships they would send those Animoji to from most to least natural:

CF > SO > S > OR > NFF > NMF > TA

Except for the relative order of CF and SO, this ranking corresponds to the scale from most to least intimate posited in Herring et al. (2020c) and reproduced in the ordering of relationships in Table 1. In the remaining tables, CF is listed after S to more closely reflect this study's findings.

Based on the findings summarized above, we generated two hypotheses regarding the speech of the participants in the Animoji clips that they sent to each "relationship":

H1: If it is performed at all, speech will be performed to a greater extent with intimate relationships (SO/CF/S) than with more distant relationships (NFF/NMF/TA).

H2: There will be associations between the results for the Animoji and the results for the relationships, similar to those reported in Table 1. However, we find no a priori justification for hypothesizing a direction of influence (i.e., Animoji determines Relationship; Relationship determines Animoji).

We also generated a third hypothesis based on the findings for public Animoji clips in Herring et al. (2020a):

H3: The Animoji will exhibit different spoken performance features according to what they represent; e.g., sounds characteristics of the Animoji will more likely occur with Animoji that are known to make characteristic sounds, such as the Dog, Cat, Monkey, and Robot, than with the Rabbit, Dragon, and Poop.

Results

Disfluencies

The distribution of disfluencies is broken down by Relationship and by Animoji in Tables 2a and 2b, respectively. In all tables, the highest values are highlighted in orange, and the lowest values are highlighted in blue. The percentages for each feature and for the features combined are calculated out of the total utterances for each relationship or Animoji.

	Pause Filler	Pause	Question Intonation	Total Utterances
SO	7 (10%)	24 (34%)	6 (8%)	71 (17%)
S	19 (22%)	24 (28%)	9 (11%)	85 (20%)
CF	16 (19%)	27 (32%)	15 (18%)	84 (23%)
OR	23 (26%)	38 (43%)	12 (14%)	88 (28%)
NFF	24 (31%)	34 (44%)	16 (21%)	78 (32%)
NMF	18 (22%)	35 (43%)	17 (21%)	82 (28%)
TA	26 (35%)	27 (36%)	*14 (19%)*	74 (30%)
Total	133 (24%)	209 (37%)	90 (16%)	562 (26%)

	Pause Filler	Pause	Question Intonation	Total Utterances
🐶	13 (18%)	26 (36%)	11 (15%)	73 (23%)
🐵	16 (20%)	33 (40%)	13 (16%)	82 (25%)
🐱	17 (21%)	29 (36%)	18 (22%)	79 (27%)
🐰	23 (29%)	25 (32%)	13 (16%)	79 (26%)
💩	17 (21%)	24 (30%)	10 (12%)	81 (21%)
🐲	21 (24%)	45 (52%)	19 (22%)	87 (33%)
🤖	26 (32%)	27 (33%)	6 (7%)	81 (24%)
Total	133 (24%)	209 (37%)	90 (16%)	562 (26%)

Tables 2a and 2b. Disfluencies

Overall, 26% of utterances contain some form of disfluency. Contrary to H1, disfluencies are *least* common in messages sent to intimates. They occur most in messages sent to new female friends, new male friends, and TA's – the more distant, less natural relationships to send Animoji to, according to the study participants.

This distribution across relationships explains the distribution of disfluencies across Animoji (refer to Table 1). Frequent use of pause fillers with the Robot corresponds to frequent use of pause fillers with the TA; frequent pauses with the Dragon map on to frequent pauses with NMF and OR, and frequent pauses for Monkey can be explained by frequent pauses for NFF. For question intonation, high values for Cat and Dragon map on to high values for NFF and NMF, respectively. However, utterances addressed to the TA often use question intonation, but question intonation is infrequent with the Robot. As will be shown later, there is a reason Robot users avoid question intonation – it conflicts with monotone delivery, a key feature of Robot performances.

The following message containing multiple disfluencies was produced by a male participant sending a Monkey Animoji to a NFF. (The N in brackets stands for native speaker.)

1) "Hi:, u:m .. it was really nice to **meet** you, and .. {*fast and slurred* I just wanted to say that you seem like a} really fun **person**? and … um, if you wanna go get coffee? or something anytime soon just let me know. Ok." [P7, M, 18, N]

Prosody

Prosody refers to patterns of stress and intonation in spoken language. Tables 3a and 3b show the frequencies of lengthening, rising pitch, falling pitch, and emphasis broken down by Relationship and Animoji.

Overall, 29% of utterances contain one of the prosodic features in Tables 3a and 3b. For Relationships, the pattern roughly mirrors that for disfluencies: Marked prosody is used most in utterances addressed to intimates, especially siblings and significant others, and least in messages to more distant relationships. Close friends are the exception to this

	Length-ening	Rising Pitch	Falling Pitch	Em-phasis	Total Utter.
SO	28 (39%)	24 (34%)	13 (18%)	36 (51%)	71 (36%)
S	44 (52%)	34 (40%)	23 (27%)	32 (38%)	85 (39%)
CF	26 (31%)	15 (18%)	5 (6%)	37 (44%)	84 (25%)
OR	21 (24%)	24 (27%)	18 (20%)	33 (38%)	88 (27%)
NFF	19 (24%)	30 (38%)	12 (15%)	32 (41%)	78 (30%)
NMF	27 (33%)	15 (18%)	9 (11%)	23 (28%)	82 (23%)
TA	26 (35%)	25 (34%)	12 (16%)	17 (23%)	74 (27%)
Total	192 (34%)	170 (30%)	93 (16%)	210 (37%)	562 (29%)

	Length-ening	Rising Pitch	Falling Pitch	Em-phasis	Total Utter.
🐶	26 (36%)	23 (32%)	7 (10%)	21 (29%)	73 (26%)
🐵	34 (41%)	30 (37%)	24 (29%)	35 (43%)	82 (38%)
🐱	29 (36%)	33 (42%)	21 (26%)	33 (41%)	79 (37%)
🐰	21 (27%)	30 (38%)	17 (22%)	23 (29%)	79 (29%)
💩	36 (44%)	19 (23%)	6 (7%)	33 (41%)	81 (29%)
👹	20 (23%)	11 (13%)	9 (10%)	31 (36%)	87 (20%)
🤖	26 (32%)	22 (27%)	9 (11%)	34 (42%)	81 (28%)
Total (Avg.)	192 (34%)	168 (30%)	93 (16%)	210 (37%)	562 (29%)

Tables 3a and 3b. Prosody

pattern, however, particularly as regards pitch variations. Possible reasons for this are discussed later.

Another apparent exception is the use of rising pitch with NFF. This could be because new female friends are often addressed with the Cat, which is often performed with rising-falling-rising ('meowing') intonation. This same intonation pattern accounts for the high incidence of falling pitch for the Cat, even though it is not matched by a high frequency for NFF. Otherwise, the prosody values for the Animoji generally follow those for their associated Relationships (e.g., Poop follows SO for lengthening; Monkey follows S for falling pitch; Monkey follows SO for emphasis). However, for emphasis, the (low) value for Robot is inconsistent with the (high) value for TA – Robot users are more emphatic than their use with the TA would predict.

Finally, the use of rising pitch with Rabbit can be related to the values for NFF or S, as both relationships were often matched with the Rabbit, as shown in Table 1. In cases where messages were sent to younger siblings by the Rabbit, rising pitch may also index the addressee's relative youth.

Example 2 sent by a female participant to a sibling via the Monkey illustrates the use of marked prosody. (NN in brackets stands for non-native English speaker.)

2) "Hi Daniel. Uh I'm sending you this **monkey** because we like sending each other <**monkeys**> and I usually use it over <**facetime**> and oh look at my eyebrows. O:::h. [opens mouth] Oh my gosh. Look at my facial **expre::ssions**. <**A:::h**<" [P21, F, 18, NN]

Special Vocalizations

Special Vocalizations include sounds or accents broadly associated with an Animoji – both the types of sounds that occur and the number of words spoken in that voice; marked vocal qualities (such as singsong, vocal fry, and accents such as AAVE, Jamaican, "Old West"); and colloquial pronunciations (such as *goin', gimme,* and *youse*) and expressions

	Animoji Sounds	#Words Sounds	Vocal Quality	Colloquial	Total Utter.
SO	11 (15%)	22 (31%)	11 (15%)	7 (10%)	71 (18%)
S	18 (21%)	32 (38%)	23 (27%)	21 (25%)	85 (28%)
CF	8 (10%)	13 (15%)	12 (14%)	23 (27%)	84 (17%)
OR	8 (9%)	19 (22%)	15 (17%)	13 (15%)	88 (16%)
NFF	7 (9%)	19 (24%)	18 (23%)	17 (22%)	78 (20%)
NMF	8 (10%)	14 (17%)	23 (28%)	26 (32%)	82 (22%)
TA	11 (15%)	37 (50%)	12 (16%)	4 (5%)	74 (22%)
Total	71 (13%)	156 (28%)	114 (20%)	111 (20%)	562 (20%)

	Animoji Sounds	#Words Sounds	Vocal Quality	Colloquial	Total Utter.
🐶	19 (26%)	26 (36%)	12 (16%)	10 (14%)	73 (23%)
🐵	7 (9%)	19 (23%)	12 (15%)	11 (13%)	82 (15%)
🐱	11 (14%)	23 (29%)	9 (11%)	21 (27%)	79 (20%)
🐰	8 (10%)	16 (20%)	15 (19%)	14 (18%)	79 (17%)
💩	7 (9%)	10 (12%)	20 (25%)	22 (27%)	81 (18%)
🐲	7 (8%)	10 (11%)	30 (34%)	23 (26%)	87 (20%)
🤖	13 (16%)	52 (64%)	16 (20%)	10 (12%)	81 (28%)
Total (Avg.)	72 (13%)	156 (28%)	114 (20%)	111 (20%)	562 (20%)

Tables 4a and 4b. Special Vocalizations

(such as *Yo, bro*, and *whaddup*). The distribution of these phenomena is broken down by Relationship and by Animoji in Tables 4a and 4b.

Overall, 20% of utterances contain one of the special vocalizations in Tables 4a and 4b. Messages sent to S strongly favor these vocalizations, especially when sent via the Dog. More surprisingly, special vocalizations are favored when addressing a TA and NMF, the most socially distant relationships. Specifically, messages are often sent to the TA via the Robot in a robotic monotone, sometimes accompanied by beeping sounds. Messages are sent to a NMF via the Dragon using colloquialisms and often speaking in a low-pitched, creaky voice, sometimes with a dialect accent such as AAVE, Jamaican, or Mexican. The Poop is performed similarly, especially when addressed to a male CF.

Example 3 was sent by a male to a NMF using the Dragon. It illustrates sounds intended to be from the Animoji (i.e., growling), a Mexican accent, and colloquialisms:

3) "{*Mexican accent* >Yee what's <**up** dude} ((laughs)) Eh [sticks out tongue] Oo::h ((growls)) ((laughs)) What's **up** bro. How's it goin'. Just wanted to see how you're doin'. Er. Uhn. [opens and closes mouth rapidly] ((laughs)) I know, it's kinda weird. [smiles]" [P18, M, 19, N]

Example 4 was sent by a male participant to a SO via the Dog; it illustrates sounds conventionally associated with the Animoji (i.e., panting and barking):

4) "Arf arf. Arf ((panting)) ((laughs)) **Hey**. Ruff. {*sing-song* That's what my d:ay w:as. R:::ough.} {*whispering* r::ough.} ((inhales)) Ouf woof ((laughing)) yeah." [P4, M, 20, N]

Animoji References

In Herring et al. (2020a), we considered lexical mentions of an Animoji or its characteristics while speaking through the Animoji to be features of performance. We add to those a new feature that emerged from our interview data, meta-mention of the Animoji as a technical artifact. The distribution of these three features is shown in Tables 5a and 5b.

Animoji References are the least common features in our data; only 8% of utterances contain one. Metamentions occurred most often in messages to NFF, possibly because they were often sent via the Monkey, which is one of the most visually expressive Animoji. Participants often made faces when using the Monkey and sometimes commented on the effects produced, as in example 2 above.

	Meta-mention	Animoji Name	Animoji Charac.	Total Utterances
SO	8 (11%)	7 (10%)	1 (1%)	71 (8%)
S	9 (11%)	9 (11%)	3 (4%)	85 (8%)
CF	6 (7%)	12 (14%)	2 (2%)	84 (8%)
OR	10 (11%)	6 (7%)	7 (8%)	88 (9%)
NFF	10 (13%)	7 (9%)	8 (10%)	78 (11%)
NMF	6 (7%)	6 (7%)	4 (5%)	82 (7%)
TA	6 (8%)	6 (8%)	2 (3%)	74 (6%)
Total	55 (10%)	53 (9%)	27 (5%)	562 (8%)

	Meta-mention	Animoji Name	Animoji Charac.	Total Utterances
🐶	10 (14%)	9 (12%)	6 (8%)	73 (11%)
🐵	14 (17%)	9 (11%)	1 (1%)	82 (10%)
🐰	2 (3%)	5 (6%)	3 (4%)	79 (4%)
🐱	8 (10%)	6 (8%)	8 (10%)	79 (9%)
💩	6 (7%)	11 (14%)	1 (1%)	81 (7%)
🐲	7 (8%)	6 (7%)	6 (7%)	87 (7%)
🤖	8 (10%)	7 (9%)	2 (2%)	81 (7%)
Total	55 (10%)	53 (9%)	27 (5%)	562 (8%)

Tables 5a and 5b. Animoji References

The higher frequency of mentioning the Animoji name with CF and S can also be explained in terms of the Animoji used most for those relationships: the Poop. Participants found this Animoji especially amusing and often referenced "poop" or its synonyms "shit" or "crap" in their messages to intimate relationships, which are the only relationships to whom it is appropriate to send the Poop, according to the study participants (Herring et al. 2020c). This example of naming is from a Poop message sent by a female to a CF:

5) "[wiggles eyebrows] Hi Lauren. ((laughs)) I'm a {*laughing* <poop emoji::<.} [opens mouth wide] Hi. U::m Let me know what you're doing after class so we can go get .. some foo::d. Okay? Bye." [P21, F, 18, NN]

Characteristics of what the Animoji represent were mentioned least often. They most commonly occurred when sending the Cat to a NFF, as in the following message to a NFF sent by a female participant:

6) "Hi. {*vocal fry* Um.} It was great **meeting** you. And {*vocal fry* um.} I'm looking forward to seeing you again. I'm sending you a cat cuz it's .. kind of feminine? I don't know. {*laughing* um but} <**yeah**< It's- it's cute. Not as cute as other Animojis but it's quite cute. So. [sticks out tongue] <**MEOW**< {*laughing* okay I shouldn't have done that.} But yeah." [P16, F, 23, N]

After bursting out with a loud "meow!", the participant semi-apologized. In so doing, she acknowledged that such playful behavior might be inappropriate in the interview setting. The implication is that she meowed in spite of herself, due to the influence of the Cat Animoji.

Discussion

In H1, we predicted that **when sending Animoji messages, speech would be performed to a greater extent in intimate relationships (CF/S/SO) than in more distant relationships (NFF/NMF/TA)**. The results for this hypothesis are mixed. H1 was SUPPORTED by the results for prosody, except for CF, who received less marked prosody than expected, given their degree of intimacy. H1 was also SUPPORTED for S for Special Vocalizations, but it was contradicted for NMF and TA. Rather than being less performed, messages sent to those relationships were performed in different ways, via marked vocal qualities and colloquialisms for NMF and (robotic) Animoji sounds for the TA. H1 was also systematically contradicted for disfluencies, especially for NFF and TA at the high end and SO and S at the low end. Disfluencies appear not to function as performance features in these data; rather, participants' speech was more disfluent when addressing relationships that they said (Herring et al. 2020c) were unnatural to send Animoji to. Finally, H1 was NOT SUPPORTED for Animoji References. Properties of the Animoji, rather than of relationships, best explained references to the Animoji and their characteristics.

H2 predicted that **there would be associations between the results for the Animoji and the results for the relationships, similar to those reported in Table 1**. This hypothesis was broadly SUPPORTED. Moreover, the association was directional, in that the Relationship results largely predicted the Animoji results for Prosody and Special Vocalizations (and for Disfluencies, in the opposite sense). A few exceptions to the associations in Table 1 were noted, however: 1) Question intonation was not associated with the Robot, but it was used most often with the TA, where tentativeness serves to mitigate the face threat inherent in messaging an instructor. 2) Conversely, emphasis was used with Robot, but not for the TA, again presumably to avoid face threat. 3) Falling pitch was more frequent for Cat than for NFF, due to a tendency for the Cat to be voiced with a "meowing" quality. Finally, 4) Rabbit performances were more colloquial than predicted by OR, NFF, or S, due to the Rabbit's association with the cartoon character Bugs Bunny.

The opposite direction of association was found for Animoji References, where the Animoji better explained the Relationship results than vice versa. Monkey and Dog seem to encourage metamentions, due to being among the most expressive Animoji; their distribution helps explain higher metamentioning with NFF, OR, S (ex. 2), and SO. Poop appears to drive Animoji naming, especially with CF (ex. 5). Finally, Cat was most often the subject of references to Animoji characteristics, which could explain the high number of characteristic mentions in messages to NFF (ex. 6).

Last, in H3 we hypothesized that **the Animoji would exhibit different spoken performance features according to what they represent**; for example, sounds characteristics of the Animoji will more likely occur with Animoji that are known to make sounds. This was broadly SUPPORTED, in that Dog and Robot made the most characteristic sounds, followed by Cat; and Dragon, Poop, and Rabbit made the fewest characteristic sounds. Monkey was the only exception: Monkey sounds were infrequent in our sample, even though real Monkeys make characteristic noises.

Tables 6a and 6b summarize the features most and least associated with each Animoji and with each Relationship. The tables also provide average *vocal performativity* (VP) scores for prosody and special vocalizations combined. These two categories best capture the vocal modifications (as compared to the lexical references and disfluencies) made by the study participants and most closely approximate the 'degree of performance' measure reported in Herring et al. (2020a). Average vocal performativity was calculated by adding the number of prosody features and special vocalizations together and dividing by the number of utterances for each Animoji and each Relationship.

For Animoji, vocal performativity ranged from a low of 20.3% with the Dragon to a high of 28.1% with the Robot, although the Robot value is inflated by the high count of words spoken in a robotic voice. A relatively narrow range of performativity was also reported for publicly-posted Animoji clips by Herring et al. (2020a).

	Most Frequent Features	Least Frequent Features	Animoji Sounds	Marked Vocal Qualities	VP
🐶	Animoji sounds; # words of sounds; *Metamentions; Animoji name; Animoji characteristics*	Emphasis; *Pause Fillers*	arf arf; ruff; woof woof; grr; ((panting))	sing-song; high pitched; whispered	24.7%
🐵	Emphasis; Falling pitch; Lengthening; *Pauses; Metamentions*	Colloquial; Animoji sounds; *Animoji characteristics*	ooh-ooh-ooh; hihihihi	rapid; quiet; Jamaican accent/AAVE	26.2%
🐰	Colloquial; Rising pitch	Emphasis; Marked vocal quality; *Animoji name; Metamentions*	((audible chewing)); {voiceless AE::H}	Bugs Bunny/Elmer Fudd voice; quiet	24.5%
🐱	Rising pitch; Falling pitch; *Question intonation; Animoji characteristics*	--	meow; MEWW	vocal fry; sing-song; quiet	26.7%
💩	Lengthening; Colloquial; Marked vocal quality; *Animoji name*	# words of sounds; Animoji sounds; Falling pitch; *Pauses; Animoji characteristics*	((farting noise))	vocal fry; low pitched; laughing	23.6%
👹	Marked vocal quality; Colloquial; *Pauses; Question intonation*	Lengthening; Rising pitch; # words of sounds; Animoji sounds	RA:::H; ((growl))	vocal fry; low pitched; clipped; rapid/monotone/AAVE	20.3%
🤖	# words of sounds; Animoji sounds; Emphasis; *Pause fillers*	Colloquial; *Question intonation*	beep beep; voo:h; eh-eh-eh-eh-eh-eh (robotic laugh); ((monotone))	monotone; vocal fry; sing-song	28.1%
Italics indicate disfluencies and Animoji references. Lighter font indicates lower frequency. Marked vocal qualities are listed most frequent first.					24.8%

	Most Frequent Features	Least Frequent Features	Marked Vocal Qualities	VP
SO	Emphasis; Lengthening	Marked vocal quality; Colloquial; *Pause fillers; Question intonation; Animoji characteristics*	sing-song; rapid	26.8%
S	Animoji sounds; # words of sounds; Marked vocal quality; Lengthening; Rising pitch; Falling pitch; *Animoji name*	*Pause; Question intonation*	quiet; vocal fry; growling	33.4%
CF	Colloquial; Emphasis; *Animoji name*	Animoji sounds; # words of sounds; Marked vocal quality; Raising pitch; Falling pitch; *Pauses; Metamentions; Animoji characteristics*	vocal fry; low pitched; laughing	20.7%
OR	Falling pitch; *Pauses*	Animoji sounds; Lengthening; *Animoji name*	sing-song; soft; high pitched	21.4%
NFF	Rising pitch; *Pauses; Question intonation; Pause fillers; Metamentions; Animoji characteristics*	Animoji sounds; Lengthening	sing-song; high pitched; laughing; vocal fry	24.7%
NMF	Colloquial; Marked vocal quality; *Question intonation; Pauses*	Animoji sounds; Rising pitch; Falling pitch; *Metamentions; Animoji name*	vocal fry; low pitched; clipped; monotone; Mexican accent/AAVE	22.1%
TA	# words of sounds; *Pause fillers; Question intonation*	Colloquial; Emphasis; *Metamentions; Animoji name*	monotone; vocal fry; low pitched	24.3%
Italics indicate disfluencies and Animoji references. Lighter font indicates lower frequency. Marked vocal qualities are listed most frequent first.				24.8%

Tables 6a and 6b. Summary of Performance Features and Average Vocal Performativity (VP)

Relationship types show a broader range of vocal performativity. Messages to siblings were the most performed (33.4%), and messages to close friends were (paradoxically) the least performed (20.7%). Relationships also differ in the number and types of performance features on which they scored high and low (as highlighted in orange and blue in Tables 2-5). Messages to S score high on six performance features and score low only for disfluencies (which is a positive indicator of fluency and hence, potentially, of performance). SO was the second-most performed relationship (26.8%), with two high scores (for prosody) as well as low scores for disfluency. But SO scores low on three other performance features, which lowers its average. CF actually scores high on three performance features and low on one

disfluency measure, but CF also has the lowest scores for seven performance measures; this accounts for its low overall score. The next lowest score, for older relatives, reflects that OR is high for only one performance feature (falling pitch) and low for three. The TA also has one high performance feature (number of words in Animoji sounds), which is counterbalanced by four low scores and a high number of disfluencies. NFF and NMF also have high numbers of disfluencies, but performance is avoided more for NMF than for NFF. NFF have three high and two low performance scores, compared to two high and five low scores for NMF.

Thus S (especially) and SO support the broader association of performance with intimacy, but NFF, NMF, and TA have higher vocal performativity scores than expected, and the low score for CF clearly contradicts the hypothesized association. The findings for the more socially distant relationships can be explained by the speaker feeling awkward when sending Animoji to those relationships (resulting in disfluencies) and using humor to mitigate the awkwardness. The findings for CF might be explained by the fact that the study participants reported sending Animoji to their CF most often (Herring et al. 2020c); the novelty of their use – and the Proteus Effect (Yee and Bailenson 2007) – might have worn off. However, this does not explain higher performativity with SO and S, to whom the participants also send Animoji often. An alternative explanation is that the participants' identity vis-à-vis their CF is indeed performed, but in a "cooler," less silly way (i.e., via colloquialisms, emphasis, and the crudeness of the Poop Animoji, but avoiding Animoji sounds, pitch variations, marked vocal qualities, etc.). If so, counting the number of performance features might not always be the most revealing way to assess performativity.

A final observation is that the Animoji messages sent by our study participants bear structural similarities to voicemail messages (cf. Gold 1991). The Animoji messages often followed the basic three-part structure of opening, body, and closing (exx. 1, 5, 6). Openings often provided a ritualized greeting such as 'hello' or 'hey,' and sometimes mentioned the caller's name (exx. 2 and 5). Closings were typically signaled by words like "ok" or "yeah" and falling intonation (exx. 5 and 6). The messages often contained disfluencies such as pauses and pause fillers, consistent with unplanned speech (Ochs 1979), although the distribution of these features appears to be determined by degree of relationship intimacy more than feeling awkward with the medium. Finally, humor and laughter were used, especially by males, albeit not to the extent observed by Herring et al. (2020a) for publicly posted Animoji clips. The similarities are likely due to the fact that both voicemail and Animoji clips are asynchronous, technology-mediated, "one-sided social actions" (Alvarez-Caccamo and Knoblauch 1992). Moreover, we might have inadvertently predisposed participants to follow the voicemail format by requiring verbal messages of a certain length (15-30 seconds). Comments made by some participants indicated that this was not how they normally used Animoji (Herring et al. 2020b).

Conclusion

In this study, we analyzed the speech produced by iPhone users in sending Animoji messages to seven common relationship types. We found fewer performance features (especially, fewer Animoji references and fewer Animoji sounds) than in the publicly shared Animoji clips analyzed in Herring et al. (2020a), likely because the messages were created in a lab setting as part of a "serious" task. However, there was still considerable evidence of playful verbal performance, in support of the earlier study. Indeed, this evidence is all the more compelling given the non-playful and non-naturalistic context in which it occurred.

Because Animoji are relatively under-investigated and include a vocal component, the present work provides a new perspective on factors that affect communication via graphicons. Specifically, it shows that people tend to modify their vocalizations in Animoji messages, even when they are not trying to perform (e.g., when discussing plans for meeting up with a friend), and they do so differently with different Animoji. This supports the Proteus Effect, or the tendency for people to conform to expectations of their digital representations (Yee and Bailenson 2007). Moreover, this is the first study to analyze Animoji use in private messages (versus users' self-reported preferences, as in Herring et al. 2020b) and that takes into consideration the sender's relationship with the message recipient. The findings show that performance via Animoji indexes not just the performer's identity (cf. Bauman 1975) but also the degree of social distance with the addressee. At the same time, even though participants reported that a CF was most natural and a TA least natural to send Animoji to (Herring et al. 2020c), the vocal performativity scores for these two relationships are reversed, indicating that intimacy alone does not fully explain performativity. 'Awkwardness' or 'coolness' can also be motivating factors. Last, we identified previously unnoticed similarities between Animoji messages and voicemail.

A limitation of this study is its small sample size. It should be repeated on a larger scale to enable statistical analyses such as summary correlations. Moreover, the study was conducted in a laboratory setting, provided limited choices of Animoji and relationship types which participants were required to match, and required participants to send messages using words rather than just facial expressions. Thus our results might not generalize to contexts of authentic use involving different Animoji (or Memoji) and other relationships. Future research should examine the messages of diverse participants, as well as naturally-occurring Animoji messages, with different Animoji and different kinds of addressees. Although we did not break down senders and recipients by gender, findings from Herring et al. (2020a,b) suggest that this would be a fruitful direction for future research. Individual differences in Animoji vocalizations should also be investigated. Finally, we recommend a deeper analysis of the structure and content of Animoji mes-

sages compared with audio, video, and text messages on mobile phones. The parallels with voicemail could provide a useful point of departure for such work.

References

Alvarez-Caccamo, C., and Knoblauch, H. 1992. 'l was calling you': Communicative Patterns in Leaving a Message on an Answering Machine. *Text & Talk* 12(4): 473-506.

Bauman, R. 1975. Verbal Art as Performance. *American Anthropologist* 77(2): 290-11.

Coupland, N. 2001. Dialect Stylization in Radio Talk. *Language in Society* 30(3): 345-75.

Gold, R. 1991. Answering Machine Talk. *Discourse Processes* 14(2): 243-60.

Haas, G.; Gugenheimer, J.; Rixen, J. O.; Schaub, F.; and Rukzio, E. 2020. "They Like to Hear My Voice": Exploring Usage Behavior in Speech-Based Mobile Instant Messaging. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1-10. New York, NY: ACM.

Herring, S. C., and Dainas, A. R. 2017. "Nice picture comment!" Graphicons in Facebook Comment Threads. In *Proceedings of HICSS-50.* Los Alamitos, CA: IEEE.

Herring, S. C.; Dainas, A. R.; Lopez Long, H.; and Tang, Y. 2020a. Animoji Performances: "Cuz I can be a sexy poop." *Language@Internet* 18: article 1.

Herring, S. C.; Dainas, A. R.; Lopez Long, H.; and Tang, Y. 2020b. Animoji Adoption and Use: Gender Associations with an Emergent Technology. In *Proceedings of Emoji2020*. Menlo Park, CA: AAAI Press.

Herring, S. C.; Dainas, A. R.; Lopez Long, H.; and Tang, Y. 2020c. "If I'm close with them, it wouldn't be weird": Social Distance and Animoji Use. In *HCI International 2020 Proceedings: Late Breaking Work*. Springer.

Konrad, A.; Herring, S. C.; and Choi, D. 2020. Sticker and Emoji Use in Facebook Messenger: Implications for Graphicon Change. *Journal of Computer-Mediated Communication* 25(3): 217-35.

Janssen, J. H.; Ijsselsteijn, W. A.; and Westerink, J. H. 2014. How Affective Technologies can Influence Intimate Interactions and Improve Social Connectedness. *International Journal of Human-Computer Studies* 72(1): 33-43.

Jefferson, G. 2004. Glossary of Transcript Symbols with an Introduction. In *Conversation Analysis: Studies from the First Generation,* edited by G. H. Lerner, 13-31. Amsterdam: John Benjamins.

Lee, J. Y.; Hong, N.; Kim, S.; Oh, J.; and Lee, J. 2016. Smiley Face: Why We Use Emoticon Stickers in Mobile Messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 760–66. New York, NY: ACM.

Linneman, T. J. 2012. Gender in Jeopardy!: Intonation Variation on a Television Game Show. *Gender and Society* 27(1): 82-105.

Ochs, E. 1979. Planned and Unplanned Discourse. In *Discourse and syntax*. 51-80. Leiden, Netherlands: Brill.

Pazil, N. H. A. 2018. Face, Voice and Intimacy in Long-distance Close Friendships. *International Journal of Asian Social Science* 8(11): 938-47.

Prager, K. J. 2000. Intimacy in Personal Relationships. In *Close Relationships: A Sourcebook*, edited by C. Hendrick and S. S. Hendrick, 229–42. New York: Sage.

Reis, H. T., and Shaver, P. 1988. Intimacy as an Interpersonal Process. In *Handbook of Personal Relationships: Theory, Research and Interventions*, edited by S. Duck; D. F. Hay; S. E. Hobfoll; W. Ickes; and B. M. Montgomery, 367–89. Hoboken, NJ: John Wiley & Sons.

Wang, S. S. 2016. More than Words? The Effect of Line Character Sticker Use on Intimacy in the Mobile Communication Environment. *Social Science Computer Review* 34(4): 456–78.

Yee, N., and Bailenson, J. 2007. The Proteus Effect: The Effect of Transformed Self-representation on Behavior. *Human Communication Research* 33(3): 271-90.

Appendix: Transcription Conventions for Animoji Clips

(( ))	Nonlinguistic vocalization ((laugh)), ((exhale)), ((panting))
{ }	Manner in which words are spoken, e.g., qualities such as vocal fry, breathiness. Enclose speech in { }. {*sing-song* Hey Gram}
[ ]	Nonverbal behavior [winks], [opens mouth], [sticks out tongue]
/ /	Transcriber uncertainty. One ? per unheard syllable. I am excited about /the/ upcoming /??/.
… / ..	Pauses: .. for 2 seconds, … for 3 seconds u:m .. it was really nice to meet you, and … I just wanted to say that you seem really fun?
(#)	Timed pause of more than 3 seconds I'd say (5) fifty-thousand
um/uh/er	Pause fillers. Write out and count as individual words.
: or ::	Lengthening of a preceding sound So:: I:'ll see you:: … next week.
<	Rising pitch. Enclose raised pitch over a stretch of speech between two <. I'm a <poop emoji<.
>	Falling pitch. Enclose lowered pitch over a stretch of speech between two >. Hey sis, how are >you::,
Bold	Emphasis; verbal stress and she knocked my water bottle **off the table**.
CAPS	Loud Monkey <SEE, monkey <DO>
*	Quiet *Kay. *Bye.
.	Utterance-final falling intonation, conventionally represented by a period in writing I spent 6 hours on tik-tok this week. Disgusting.
,	A slight pause indicating continuation, conventionally represented by a comma in writing It's pretty embarrassing, I can't believe it but,
?	Questioning intonation, independent of question syntax; can appear mid-sentence (cf. "uptalk") I have a question? .. about the assignment?
-	Break, shift in thought, self-interruption ("false start") I know you said that you log- like a dog before,