Are you laughing, smiling or crying?

Donna Erickson*, Caroline Menezes† and Ken-ichi Sakakibara††
*Showa Music University, Kawasaki 251-8558 Kanagawa Japan
E-mail: EricksonDonna2000@gmail.com Tel: +81-90-8725-7412
†University of Toledo, Toledo, Ohio, U.S.A
E-mail: cmeneze@utnet.utoledo.edu
††Health Sciences University of Hokkaido, Sapporo, Japan
E-mail: kis@hoku-iryo-u.ac.jp

***Abstract*— Acoustic, articulatory, and perceptual analyses of spontaneous laughing, smiling, and crying speech were done in comparison with neutral speech. Listeners were asked to rate the emotional intensity and identify the emotion as *happy*, *sad*, or *neutral* (or other/unknown) of auditorily presented (a) phrases and (b) single words. The results show acoustic, articulatory and perceptual similarities for laughing, smiling and crying speech; smiling speech was sometimes judged as *sad*. Utterances rated as emotionally intense (whether laughing, smiling, or crying speech) are characterized by high F0, high F2 and low H2 (dB) (especially for *happy*), and tended to be produced with raised/retracted upper lip, and lowered tongue dorsum. Possible reasons for the phonetic similarities in such divergent types of emotional expressions, e.g., laughing, smiling and crying, are discussed. Also, discussed are possible reasons why phonetic characteristics of speech intended by the speaker to be emotional are different from those perceived by listeners.**

I. INTRODUCTION

Laughing, smiling and crying occur frequently in human expression of emotion. These emotions can occur simultaneously with speech. In speech laughs, vowels and consonants of speech co-occur with the amplitude contour, breathiness, and rhythm of laughter. Kohler [1] referred to speech laugh as "free laughter superimposed on speech", characterized by several instances of rhythmical energy outputs during speaking the word. According to Nwoka et al. and Trouvain [2, 3], 50% of laughter production consists of speech laughs.

Increased F0 is reported for laugh and speech laugh compared to speech ([3,4]. Menezes and Igarashii [5] in addition show a gradual increase in the amplitudes of higher harmonics (decreasing values of H1-A3) as well as a progression of lowering of formant frequencies, going from speech to speech laugh to laugh. Increased H1-H2 and H1-A3 may be associated with an increase in breathy voice quality (e.g., [6]) as one goes from speech to speech laugh to laugh.

In addition to speech laugh, there are 'speech smiles', used to vocally express happiness, accompanied by high pitch and facial expression of happiness, e.g., smiling [7, 8]. Speech smiles contrast with speech laughs in that the breath control of speech is used [1]. They are generally recognized auditorily as smiled speech because of lip spreading with smiling, which may lead to increases in F2 and F3 [9, 10, 11]. For speech smile, it is said that the mouth is open with lip corners pulled upward and backward [12], the tongue may be fronted [1] and the vocal tract may be shortened due to increased mouth opening, along with a coupling of laryngeal tension with facial tension [2, p. 882] and possible raised larynx [13]. For speech laugh, the tongue may be relaxed, but not the jaws and lips [14]. The jaw may be lower with tongue dorsum retracted to account for lower formant frequencies [4]. Darwin [15] observed for laughter a low jaw, open mouth with corners retracted and slightly raised, and upper lip also raised, which Darwin, together with Duchenne, postulated is caused by both the zygomatic muscles of the face and the obicularis muscles of the eye. Interestingly, Darwin observed that the facial description for laugh is very similar to that of cry, and along these lines, DeBenedictus [16] reported that laughing sounds can be frequently confused with cries. Titze et al. [17], based on modeling studies together with EMG studies, suggest that lung pressure/lung volume define the number of bursts/calls in a giggle bout but that the laryngeal muscles exert the primary control of voice fundamental frequency.

Esling [18] reported an example of laughter which used a low larynx but high pitch and the supraglottic space (often referred to in some literature as hypopharyngeal area) is not constricted but open. A similar finding was reported by Estill [19] in her training manual for the singing voice: a low larynx along with a wide, non-constricted aryepiglottic region, produces a voice quality characteristic of both cry and laugh (p.111), regardless of whether the F0 is high or low. Acoustic consequences of lowered larynx together with expanded hypopharyngeal area are decreased F2, F3, F4 and a trough of energy around 5 khz (e.g., 20, 21, 22). Erickson et al. [23] reported lowered F2, F3, and F4 for *sad* (crying) speech (on the vowel /i/).

In this paper we examine the acoustic, articulatory and perceptual characteristics of spontaneous laughing, smiling and crying speech as compared with spontaneous "neutral" speech. Earlier reports on a subset of this data have been reported in [24, 25].

II. METHODS

Acoustic and articulatory recordings were done using the 2D EMA system (NTT Research Laboratories, Atsugi, Japan) for an American Midwest dialect female speaker (first author) in an informal spontaneous telephone dialogue with a

conversation partner (second author) through an earphone/microphone set-up, where the conversation partner sat in a separate room from the subject. The conversation partner asked the subject various unrehearsed questions based on a list of topics related to the subject's personal life to evoke emotions. Laughing, smiling and crying speech were well-evoked, especially the latter, since the subject was at the time of the experiment grieving the loss of her mother. The dialogue continued in a natural manner, while EMA recordings were made within a window frame of 20 sec, with a break in recording of about 3 seconds between frames. Acoustic recording, however, was continuous. Video recording was also done but not used for systematic analysis. From this large data base, a small subset was selected for perceptual analysis: 39 phrases containing the same or similar words, spoken while the speaker was laughing, smiling, crying or talking in a non-emotional manner. From these, a subset of 24 words –eight sets of triplet words (*happy*, *sad*, *neutral*) --were selected for further perceptual, articulatory and acoustic analysis.

Articulatory analysis was done by examining the movement of the EMA receiver coils attached to the (1) lower incisor (mandible) (2) upper lip, (3) lower lip, and (4) receiver coils (T1, T2, T3, T4) attached along the longitudinal sulcus of the speaker's tongue. The positions of the transmitter coils determine the coordinate system [26] with the origin positioned slightly in front of and below the chin. All EMA values are positive, with increasingly positive y-values indicating increasingly raised lip, jaw or tongue position, and increasingly positive x-values, increasingly retracted lip, jaw or tongue position.

Articulatory measurements were made for the x-y coil positions for the upper and lower lip (UL, LL), for the mandible (J), and the tongue (T1, T2, T3), using a MATLAB-based analysis program. Articulatory and acoustic measurements were made at the time of maximum jaw opening for each of the words analyzed. A sample of the point of maximum jaw opening where the acoustic and articulatory measurements were made is shown in Figure 1, which compares *happy* (smiling), *sad* and *neutral* speech.

Acoustic measurements of duration and average F0 were made for each word analzyed. In addition, F0, F1, F2, F3, F4 and voice quality measurements (H1-H2, H1-A3) were made at the point of maximum jaw opening. As mentioned in section 1, H1 is the amplitude (dB) of the first harmonic (F0), H2, of the second harmonic, and A3, of the strongest harmonic associated with the third formant. Increased values of H1-H2 and H1-A3 are said to be indicative of voice quality associated with breathiness. Intensity was not measured, because the microphone placement was not constant during the course of the experiment. The corpus contains an unbalanced mixture of vowel types: 4 high front, 3 mid front, 6 low front, and 11 rounded mid back vowels. Detailed acoustic and articulatory analyses reported in this paper focus on the /o/ vowel utterances. Only 10 items were used for acoustic analysis because "going" (*sad*) was spoken as a "sob" and only 9 were used for articulatory analysis because the articulatory recording of "don''t" (*sad*) was incomplete (the utterance was spoken at the end of the 20 s window for collecting EMA data).

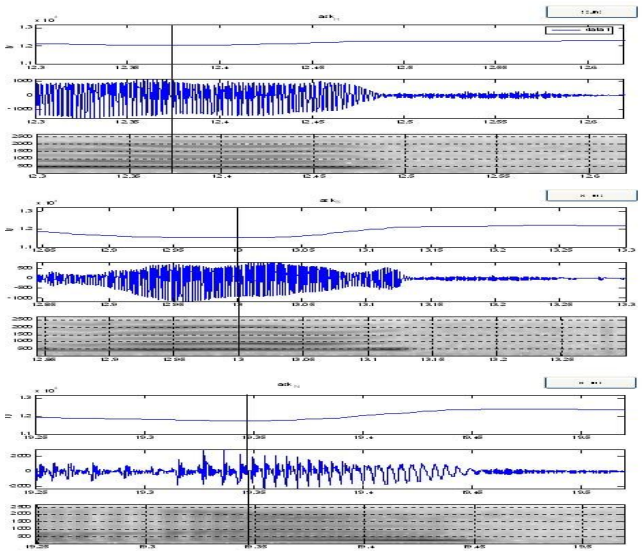

Figure 1. Sample of *happy* (smiling) speech (top) with *sad* speech (middle) with *neutral* speech (bottom) for the word "ask". The vertical line indicates the point of maximum jaw opening at which the articulatory and acoustic measurements were made. The top window in each panel represents the jaw-y movement, with a range of 1.1 to 1.3 mm, the middle window, the acoustic signal, and the bottom window, the spectrogram. The x-axis shows time in ms. Tongue dorsum is not displayed.

The downloadable free acoustic software WaveSurfer and Praat were used for the acoustic analysis. For the *laugh* speech, more than one instance of acoustic/ energy chunks was found. In order to compare the measurements of the *laugh* speech with that of the other categories of speech (i.e., *smile*, *sad*, and *neutral*), F0, formant and voice quality measurements were made for only the first acoustic chunk of the laugh speech.

Two perception tests were administered: a phrase test (total of 39 short phrases, 3 randomizations) and a single word test (total of 36 words, mostly monosyllabic, 3 randomizations, 27 of which were taken from the phrase tests). Perceptual, acoustic and articulatory analyses were done for 24 of these words—8 sets of triplets. Four of the triplet-sets involved smile speech (SS), and four involved laugh speech (LS). Table 1 shows a list of phrases and words used in the perception tests.

The tests were administered auditorily through HDA200 Sennheiser headphones in a quiet room, using a Windows-based computer software from Runtime Revolution. The listeners were asked to respond to two questions for each stimulus: (1) rate each word according to the perceived degree of emotion on a 5 point scale, with "5" most *emotional* and "0" *not emotional at all*; (2) identify the perceived emotion—

(1) *happy*, (2) *sad*, (3) *no emotion/neutral*, (4) *other* (5) *unknown*. The listeners could listen as often as they wished to each sound. A practice test of 5 utterances preceded the test. The listeners were 79 Midwestern American college students from Capital University (Columbus, Ohio) and Black Hills State University (Spearfish, South Dakota).

Table 1. Phrases and single words spoken with *happy*, *sad*, *neutral* emotion on different vowels used in the two perception tests. The word "going" (Sad) was so short that no vowel was perceptible.

intended emotion	word	phrase	vowel
Happy (SS)	**attention**	mom, you're not paying **attention**	ɛ
Sad	**attention**	Pay **attention**	ɛ
Neutral	**attention**	And I should have paid **attention** to the…	ɛ
Happy (SS)	**ask**	You **ask** my husband	æ
Sad	**ask**	You can **ask** for help	æ
Neutral	**ask**	If I **ask** you questions	æ
Happy (SS)	**sad**	That makes me **sad**	æ
Sad	**sad**	I'd be very, I'd be very **sad**	æ
Neutral	**sad**	happy and **sad**	æ
Happy (SS)	**oh**	**Oh**, Happy Day!	o
Sad	**so**	**So**, I would miss that very much	o
Neutral	**so**	and **so**, something that make me feel sad	o
Happy (LS)	**going**	gotta video tape **going** 〈laugh〉	o
Sad	**going**	Like when somebody's **going** to leave you	
Neutral	**going**	think we're not **going** to get any real emotion	I
Happy (LS)	**know**	You **know** 〈laugh〉	o
Sad	**know**	you **know** it	o
Neutral	**know**	You **know** I flew out on the 12th	o
Happy (LS)	**thing**	Do such a **thing** 〈laugh〉	I
Sad	**think**	You **think**	I
Neutral	**think**	I **think** so too	I
Happy (LS)	**don't**	**Don't** go east	o
Sad	**don't**	Pay attention **don't**	o
Neutral	**don't**	**don't** forgive	o

III. RESULTS

A. Perception Tests

Table 2 (see end of paper) shows listeners' categorizations and ratings of intensity of the emotion for each of the items in the perception tests (1) with single words (indicated in **bold** type in the table) and (2) with phrases in which those words occurred. Interestingly, even when listeners were presented with the phrases, there was not 100% identification of the speaker's intended emotion. But there was above chance level of identification (e.g., above 20%, since there were 5 choices--*happy*, *sad*, *neutral*, unknown, other). *Sad* utterances were best perceived as *sad* (76%) while *happy* were perceived at 58% and *neutral* at 47%. Those phrases that were well-identified as either *happy* or *sad* had an average intensity rating of 3.2; those identified as *neutral*, 1.8.

The identification of emotion for the phrases was sometimes influenced by the semantics of the utterance; for instance, utterance 6 ("That makes me sad") spoken as *happy* but with a lexical meaning of *sad* was only identified 8% of the time as *happy*, and utterance 24 ("don't forgive") spoken as *neutral* but having a negative morpheme was identified as *neutral* only 30% of the time. In addition, certain *happy* phrase utterances even with a *neutral* meaning were sometimes perceived as *sad*: Utterances 7 and 8 ("Mom, you're not paying attention" and "You ask my husband") were heard only 9% and 5%, respectively, of the time as *happy* and 53% and 69% of the time as *sad*. Both of these utterances were spoken with a smile (according to the video recording), but nevertheless were heard predominantly as *sad*, not *happy*.

The perceptual results with single words were to a certain extent similar to that for the phrase utterances. Emotion can be identified by listeners in single words well above chance. Intended *sad* words were perceived as *sad* 52%, *happy* as *happy* 56% and *neutral* as *neutral* as 44%. In certain cases, the identification of emotion of the word was better than that of the phrase, e.g., for utterances 11 ("pay attention **don't**"), and 12 ("you **think**"), both spoken as *sad*, the single words (**don't** and **thin**k) were better identified as *sad* than the phrases (76% for the single word vs. 44% for the phrase, and 76% for the single word vs. 13% for the phrase, respectively). Also, the opposite was seen: words intended to be *sad* were sometimes perceived as *neutral*. For instance, utterances 13-16 ("so," "know," "going," "don't") spoken as *sad* were heard as *neutral* (40%, 47%, 64%, 52%, respectively).

Also, as with the phrases, those words that were well-identified as either *happy* or *sad* had an average intensity rating of 3.2; those identified as *neutral*, 1.8

N*eutral* speech was never given an intensity rating of 3 or above, but sometimes *sad* speech was rated as *neutral*. *Happy* laugh speech utterances ("thing," "going," "know," "don't") were well-recognized by listeners as *happy* speech, i.e., 83%, 81%, 84% and 95%, respectively. *Happy* smile speech utterances ("attention," "ask," "sad," "oh") were not that well recognized as *happy* speech, i.e., 18%, 18%, 17%, 48%, respectively, and sometimes were identified as *sad*, e.g., *happy* smile speech utterances, "attention," "ask," and "sad" were identified as *sad* utterances, i.e., 48%, 57%, 41%, respectively.

The results with phrases and single word utterances suggest that (1) listeners generally can identify the intended emotion of the speaker by listening to a phrase or word out of context of the entire conversation, (2) that utterances that were identified as emotional also were assigned relatively high emotional intensity ratings, and those identified as *neutral*, low intensity ratings, and (3) the speaker's intentions and the listener's perceptions do not match 100% of the time.

B. Single words. Articulatory/acoustic measurements

Acoustic measurements for all single word utterances with the vowel /o/ are discussed here (Table 2a). According to the table, *happy* and *sad* words, compared to *neutral* words, group together in terms of F0 and formant frequencies: F0 tends to be higher, F1 lower, F2 higher, and F4 lower. In terms of amplitude of harmonics, *happy* seems to be different from *sad* and *neutral*: H1(dB), H2(dB) and H1-A3(db) are smaller, but H1-H2 is larger. ANOVA with the acoustic measures as dependent variables and "intended emotion" as the independent variable found main effects only for H2 (p=0.002). T-tests show that *happy* vs *neutral* speech is significantly different in terms of average F0 (p=0.046) and H2 (p=0.001) with *happy* speech having higher F0 and lower H2 (dB) than *neutral*, but there are no significant differences between *happy* vs *sad* speech (p=0.258, p=0.173, respectively). To summarize, *happy* and *sad* speech have higher F0 than *neutra*l; *happy* and *sad* speech have lower H2 than *neutral*; *happy* has larger H2-H1 than *sad* or *neutral*; and *happy* has smaller H1-A3 than *sad* or *neutral*.

Table 2b shows the mean articulatory values for *happy*, *sa*d, and *neutral* utterances for words on the vowel /o/. Lower values of y measurements indicate lower articulator position, i.e., lower jaw, lip and tongue. Lower values of x-measurements indicate more protruded/advanced articulator position, i.e., more protruded jaw, lip and tongue. A general tendency seen from the mean values in Table 2b is that *happy* and *sad* /o/-vowel utterances compared with *neutral* ones have lower jaw and more retracted upper lip. In addition, *happy* utterances have more raised upper lip, lowered/retracted lower lip, as well as lower, retracted lower lip and lower, retracted tongue dorsum.

Table 2a. Mean acoustic values of the 10 /o/-vowel utterances intended by speaker to be *happy*, *sad* or *neutral.*

Emotion	DUR (ms)	AVF0 (Hz)	F0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	H1 (dB)	H2 (dB)	A3 (dB)	H1 H2 (dB	H1 A3 (dB)
H (av of 4)	0.41	298	301	365	1662	2865	3996	-43	-51	-58	8.2	15.8
S (av of 3)	0.36	243	239	375	1673	2725	3908	-36	-41	-69	4.9	33.2
N (av of 3)	0.31	190	197	410	1541	2992	4125	-32	-37	-65	5.2	33.7

Table 2b. Mean articulatory values of 9 /o/-vowel utterances intended by speaker to be *happy*, *sad* or *neutral.*

Emotion	JX	JY	ULX	ULY	LLX	LLY	T3X	T3Y
H (av of 4)	7.2	12.0	6.5	14.5	6.5	13.0	10.8	14.1
S (av of 2)	7.3	12.0	6.5	14.3	6.2	13.3	10.4	14.5
N (av of 3)	7.4	12.3	6.3	14.2	6.3	13.3	10.6	14.7

An ANOVA done for the utterances on the vowel /o/ with "intended to be emotional (e.g., *happy/sad*)" vs. "not-intended to be emotional (e.g., *neutral*)" as the independent variable and the articulatory measures as the dependent variables found main effects only for upper lip-x (p=0.002). T-tests show that upper lip-x is significantly different for *happy* (laugh and smile) speech vs. *neutral* speech (p=0.009) with *happy* speech more retracted than for *neutral* speech, but there are no significant differences between *happy* and *sad* speech (p=0.627).

To summarize the acoustic and articulatory characteristics of speech intended by the speaker to be emotional, ANOVA shows that *happy* (laugh and smile) speech (as well as *sad*, including crying speech), as opposed to *neutral* speech, is characterized by high F0, low H2, and retracted upper lip. It is interesting that t-tests show no significant differences between *happy* and *sad* speech in terms of their acoustic and articulatory characteristics.

Although laugh and smile speech are similar in terms of acoustic and articulatory measurements, laugh speech is unique, in that it is simultaneously speech and laugh. The top panel of Figure 2 shows a typical example of "laugh speech", as spoken on the utterance "think"(top panel), and contrasts with "thing" spoken as *sad* and *neutral*, shown in the middle and bottom panels, respectively. For laugh speech the vowel is elongated and chunked into 3 acoustic and energy events, a pattern not seen in the *sad* or *neutral* utterances. For each utterance, there is one opening-closing of the jaw (not shown here). Even for laugh speech, where the acoustic signal is divided into subparts of energy rises and falls, there is only one jaw opening. This pattern of multiple acoustic/energy chunks but one jaw opening per linguistic monosyllabic word was seen for each of the other three laugh utterances in the data and is a new finding previously not reported.

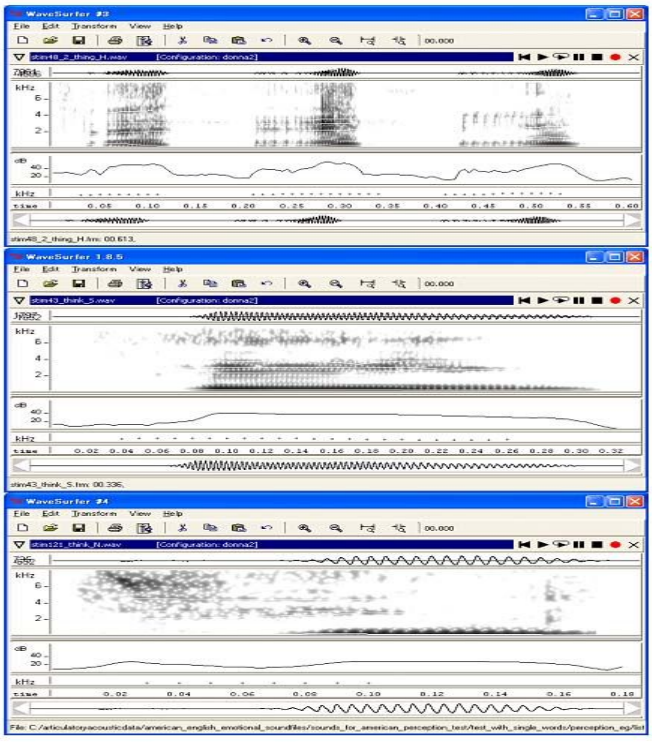

Fig. 2. The utterance "think" spoken as laugh speech (top), "thing" as *sad* speech (middle) and "thing" as *neutral* speech (bottom). Shown for each utterance is the acoustic signal, the

spectrogram, the intensity contour (dB), F0 contour (kHz), time signature, and acoustic signal.

C. Single words: perceptual/acoustic/articulatory

In the above section, we examined the acoustic and articulatory characteristics of the /o/ words *intended* to be *happy*, *sad*, or *neutral*. In this section, we examine the characteristics of the /o/ words that were *perceived* by listeners to be emotional, where "perceived to be emotional" was "1" if the averaged rating (answer to question 1) was "3" or above, and "0" if it was less than "3". Due to small sample size, it is not possible to analyze *happy* vs. *sad* speech separately. ANOVA with "perceived to be emotional" as the independent variable, and the articulatory/acoustic measures for the /o/-vowel utterances the dependent variables, found main effects for the following measures: average F0 (p=0.010), F0 (p=0.025), F2 (p=0.026), H2 (p=0.000), upper lip-x (p=0.001), upper lip-y (p=0.040), and tongue dorsum-y (p=0.021). These results suggest that /o/-vowel utterances perceived to be emotional not only had high F0, but also high F2, and were characterized by retracted/raised upper lip and lower tongue dorsum positions. Comparison of the differences in phonetic parameters for /o/-vowel utterances between "intended to be emotional" (as produced by the speaker) and "perceived to be emotional" (as perceived by listeners) shown in Table 3 indicates that only H2 and upper lip x were significantly different for those utterances intended to be emotional (*happy* or *sad*), but F0, F2, H2, upper lip-x, upper lip-y, lower lip-y and tongue dorsum-y were significantly different for those /o/-vowel utterances perceived by listeners to be emotional. This suggests that for *happy* (or *sad* crying) utterances, the speaker produced H2 with a lower amplitude and retracted upper lip; however, listeners, cued into not only these two things, but also, in perceiving *happy* (or *sad* crying) utterances, they paid attention to higher F0, higher F2, and certain acoustic cues resulting from raised upper lip and lowered tongue dorsum. That different phonetic parameters may be associated differently with intended vs. perceived emotional speech is consistent with the mixed perceptual results shown in Table 1, indicating that not all utterances intended as *happy* or *sad* or *neutral* by a speaker who is experiencing intense emotion are perceived by listeners as *happy*, *sad*, or *neutral*, respectively. No significant correlations for intensity ratings and acoustic/articulatory measurements were seen. Given the small sample size in this study, no conclusive statements at this point can be made, however.

Table 3. Significant differences in phonetic parameters between "intended to be emotional (by speaker)" and "perceived to be emotional (by listeners)" for /o/-vowel utterances

Emotion	avF0	F0	F2	H2	ul-x	ul-y	T3-y
Intended				0	0.008		
Perceived	0.00	0.03	0.026	0.00	0.001	0.04	0.02

IV. Discussion

A Similarity between happy and sad speech

The perception results with phrases and with single words, asking listeners to rate the emotional intensity of the speech sounds, showed that laugh speech was always heard as emotional (rated as "3" of above), but smile speech was not. It was given a rating of "3" or above for only half of the speech utterances (total of 2 out of 4). When listeners were asked to label the emotion they heard, laugh speech was well-perceived as *happy* speech (80% of the time or more), but smile speech, less well-perceived, unless the semantic meaning of the sentence was happy, as in the phrase, "Oh happy day!" which was identified 89% of the time as happy. Otherwise, smile speech phrases were identified as *sad* above chance (40% and 60%), and smile speech words also above chance as *sad* (21% and 57%.) Since *sad* and *happy* speech were very similar in terms of acoustics and articulation (t-tests showed no significant differences), it should not be surprising perhaps that perception results showed confusion between smile and *sad* speech. This finding is compatible with previous studies that have shown similarities between laugh and cry, e.g., [18, 15, 16]. It is extremely interesting that the acoustic and articulatory characteristics of *happy* and *sad* speech overlap, thus resulting in misperception by listeners. One possible answer for this mismatch may have to do with an inherent and fundamental similarity in physiological mechanisms underlying *happy* and *sad* emotions—the purpose of both may be to bring about a type of catharsis or relaxation for the speaker; hence, relaxation of the muscles for breathing and vocalization, which might account for the greater glottal opening for *happy* and *sad* speech (i.e., smaller H2-H1). Further exploration along these lines needs to be done.

Another avenue of thinking might be that speakers do not necessarily experience any one single emotion at a single time. Emotions are complex. In the case of this experiment, the speaker was feeling very sad because of the situation about her mother, and the underlying sadness may have colored her way of speaking, even while laughing and smiling. A similar situation may be seen, for instance, in cases of chronically depressed speakers, with the underlying emotion of depression/sadness giving a long-term effect of coloring their speech expressions but with short-term effects of temporary displays of other emotions.

B Phonetic characteristics of emotional speech (intended by speaker to be emotional, perceived by listeners to be emotional)

In terms of acoustic and articulatory characteristics, ANOVA analysis showed that laugh and smile speech (and *sad* speech) compared with *neutral* speech for /o/ vowel utterances was significantly different in terms of having high F0, low H2, and retracted upper lip. In terms of listeners' perception of whether an utterance was emotional or not, ANOVA analysis showed that utterances judged as emotional

vs. not-emotional were significantly different for /o/ vowel utterances in terms of emotional utterances having high F0, raised F2, low H2, retracted/raised upper lip and lowered tongue dorsum.

High F0 for speech laugh and smile has been previously reported, as mentioned in the introduction, e.g.,[6. 2, 8, 9]. The use of EMA in this study confirms the raised and retracted upper lips previously observed for laugh, e.g., [15] and speech smile, e.g, [12. 15, 10]. However, in this study we also find raised retracted upper lip for laugh speech, as well as *sad* speech. Our finding of lowered tongue dorsum confirms the hypothesis suggested by [6] for laugh speech, and which we also found for smile speech. There is a tendency in our data (see Table 2b) for jaw position to be lower for laugh and smile speech. This was hypothesized by [6] to be the case for laugh speech, and was observed by Darwin for laugh.

Some comments about lowered F4 for emotional speech: This may be caused by vocal tract lengthening/enlarging due to a lowered larynx, as well as an expanded hypopharyngeal region (ventricular area and piriform fossa) (e.g., [27, 20, 21, 22] According to [19], the lowering of the larynx and expanding of the hypopharyngeal region would produce the voice quality heard in laugh and cry. Erickson et al.[23] reported lowered F2, F3, and F4 for *sad* (crying) speech on the vowel /i/. A lowered tongue dorsum, as found in this study, might be part of the process involved in lowering the larynx, which would lengthen the vocal tract, and among other things, reduce F4. At this point, no larynx height data is available; however, plans are underway to collect empirical data, e.g., simultaneous fiberoptics, EMG, EGG, larynx height, video and acoustic recordings, about larynx lowering and hypopharyngeal expansion during laugh and smile (as well as *sad*) compared with *neutral* (modal) phonation in order to explore these hypotheses.

About the voice quality measures, H1-H2 and H1-A3, we do not see a progression of larger H1-A3 values for laugh speech compared to *neutral* speech, as was reported by [6]. However, we do see a difference in H2 amplitude—with H2 lower for utterances that were rated as intensely emotional. ANOVA with "perceived emotional" vs. "perceived-not emotional" as the independent variable showed that H2 was significant. Especially for the laugh utterances, which were rated intensely emotional, we see low amplitude for H2.

Laugh speech is very unique, compared with smile speech or modal (*neutral)* speech. As previously reported (e.g., 1,2], the acoustic signal is divided into smaller parts, associated with energy rises and falls. But the new finding in this study is that there is only one jaw opening per word-unit, regardless of the number of smaller chunks of acoustic and amplitude units. This is interesting from the point of the articulatory organization of speech, and specifically, the articulatory organization of the syllable. According to certain hypotheses (e.g., 28, 29], the jaw is the articulatory organizer of the syllable and the magnitude of the jaw opening dictates the prosodic strength of the syllable. The finding for laugh speech of one jaw opening per monosyllabic speech unit simultaneous with multiple respiratory and laryngeal events suggests maybe a multilayered structure of emotional speech: a first layer physiological, dealing with strengthening or relaxing muscles for breathing and vocalization, and a second layer, having to do with linguistic/social structure. Our findings here may contribute toward understanding the organization of laugh speech. That is, how does a speaker combine laughter, something part of a human's vocal repertoire before speech developed, with speech, something that requires controlled coordination of the respiratory and laryngeal/supralaryngeal articulatory mechanisms, e.g., 2, 30, 6],. The question is far from answered, but perhaps this finding will contribute to an answer in the future. Clearly, more analysis of laugh speech needs to be done.

To summarize some of the articulatory and corresponding acoustic characteristics of emotional (*happy* and *sad* crying) speech (as spoken on /o/ vowels):

1. F0 is significantly higher for emotional (*happy* and *sad* crying) speech compared to *neutral* speech. It is not immediately clear why high F0 is a characteristic of intensely emotional speech.
2. F2 is higher for *happy* and *sad* crying speech; the upper lips are raised and retracted, which would account for a raised F2.
3. There is a tendency for F4 to be lower for *happy* and *sad* crying speech; the tongue dorsum is also lower, which might possibly lead to a lower larynx and consequently reduced F4 (since the vocal tract would be lengthened). This remains to be substantiated with data from physiological experiments.
4. There is a tendency for H1-H2 to be larger, especially for *happy* (laugh) speech. This suggests a larger glottal opening and more breathy quality for *happy* (laugh) speech.
5. There is a tendency for H1-A3 to be smaller for *happy* (laugh) speech. This suggests an abrupt closing of the vocal folds, which may be caused by high sub-glottal pressure, and the consequent Bernouilli force which would cause the folds to close abruptly due to the high velocity of air particles passing through the glottis.
6. H2 (dB) is significantly lower for *happy* speech. It is not clear why this is, but needs to be researched more.

D Why is there a difference between speech intended by speaker to be emotional and speech perceived by listeners to be emotional

The results suggest there is a difference in acoustic and articulatory characteristics between speech *intended* by a speaker to be emotional and speech *perceived* by listeners to be emotional. In terms of production of *happy* (or *sad* crying) utterances, the speaker produced H2 with a lower amplitude and retracted upper lip. However, in terms of perception, listeners, cued into not only these two aspects, but also, to higher F0, higher F2, and certain acoustic cues resulting from raised upper lip and lowered tongue dorsum. It is not clear why we found these differences between production and perception. The utterances in the perception tests were taken

out of a larger context, and presumably given the complete context, listeners' perception might more accurately match the speaker's intention. Nevertheless, it remains interesting that we found this difference in this experiment. One interpretation may be that production goals of a speaker while in the throes of an emotional experience may be different from those of a speaker not overwhelmed by the emotion but who is concentrating on communicating paralinguistic/affective information to a listener, such as the. situation with expressive speech in acted situations or in the usual give and take of daily communication. As such, the acoustic characteristics may be of a more stereotypical nature and of a type that listeners may be more accustomed to hearing than those of the type recorded in this particular experiment, where the speaker was experiencing strong emotion at the same time talking.

V. SUMMARY

We examined acoustic and articulatory recordings of spontaneous *sad* crying and *happy* speech (laugh and smile speech), and comparing them with *neutral* speech. We asked listeners to (1) rate the emotional intensity of the speech utterances and (2) categorize the utterances as *happy*, *sad*, or other (not-emotional, unknown, other). The sample size was small and unbalanced, so conclusive remarks cannot be made.

Nevertheless, certain interesting and consistent characteristics emerged from this study which can be summarized as follows: (1) There are distinctive acoustic, articulatory and perceptual characteristics for "emotional" vs. "not-emotional" speech; (2) There are similarities between *happy/sad* crying speech in terms of acoustics, articulation, and perception; also between laugh and smile speech; (3) Laugh speech, even only spoken as a single word, was easily recognized as *happy*, but it was not always easy for listeners to identify smile speech as *happy* (both for single words and phrases) to the extent that smile speech was sometimes confused with *sad* speech; (4) Laugh speech is characterized by several acoustic and intensity chunks during the word, yet only one jaw opening; (5) Intended *happy* speech (laugh, smile speech) (for /o/ vowel utterances) is significantly different from *neutral* speech in terms of its higher F0, lower H2 , and retracted upper lip; and (6) Utterances (spoken with /o/ vowels) perceived by listeners as emotional are significantly different from those perceived as *neutral* in terms of higher F0, higher F2, lower H2, raised/retracted upper lip, and lowered tongue. The assumption in this paper is that the findings for the /o/-vowel utterances apply to all vowels, but a larger data base is needed to verify this.

Some future questions we wish to explore are (1) how does a speaker combine laughter with speech? (2) why did laughter emerge? and (3) what are the important voice quality characteristics of crying, laughing, and smiling speech, in terms of acoustics, articulation, and perception?

Finally, we wish to explore further the differences in acoustic and articulatory parameters of intended emotional speech (as produced by the speaker) and those of perceived emotional speech (as perceived by listeners). Both acted and spontaneous emotional speech utterances are valid topics for research, but with different research goals. The focus of this paper is on spontaneous expression of intense emotion, which we believe is important for exploring the essential mechanisms underlying speech production in humans, and will form a basis for better understanding of intra-human (as well as human machine) communication. These differences in production and perception may be crucial factors in (mis)communication among humans, as well as human-machine interfaces.

ACKNOWLEDGMENT

We thank Masaaki Honda, A. Fujino and NTT Communication Science Labs, Atsugi, Japan, for allowing us to use the EMA facilities to collect the original data, and Akinori Fujino for assisting in the data-collection. Also, we thank Albert Rilliard for helping with the software for the perception tests, and the college students at Capital University and Black Hills State University. This work was supported by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007-2010):19520371 to the first author, and part of this work was supported by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

REFERENCES

[1] Kohler, K.: 'Speech-smile', 'speech-laugh', 'laughter' and their sequencing in dialogic interaction. Interdisciplinary Workshop on "The Phonetics of laughter", Saarbrucken, 4-5 August, 2007. (2007).

[2] Nwokah, E.E.; Hsu, H-C.; Davies, P.; Fogel, A.: The integration of laughter and speech in vocal communication. A dynamic systems perspective. J.Speech, Lang. Hearing Res. 42: 880-894 (1999).

[3] Bachorowski, J-A.; Smoski, M.J.; Owren, M.J.: The acoustic features of human laughter. J. acoust. Soc. Am. 110: 1581-1587 (2001).

[4] Trouvain, J.: Phonetic aspects of 'speech laughs". Proc. Conf. on Orality and Gestuality (Orage), Aix-en-Provence, 634-639 (2001).

[5] Menezes, C.**;** Maekawa, K.; Kawahara, H.:. Perception of voice quality in paralinguistic information types. In Proceedings of the 20th General meeting of the Phonetic Society of Japan, Special issue of the 80th Anniversary. Tokyo, Japan, 153-158 (2006).

[6] Menezes, C.; Igarashi, Y.: The speech laugh spectrum. Proc. Speech Production, Brazil (2006).

[7] Fujimoto, M.; Maekawa, K. Variation of phonation types due to paralinguistic information: An analysis of high-speed video images. In 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2401-04. (2003).

[8] Tartter, V.C.: Happy talk: Perceptual and acoustic effects of smiling on speech. Percept Psychophys. 27:24-27 (1980).\

[9]Robson, J.; MackenzieBeck, J.: Hearing smiles – Perceptual, acoustic and production aspects of labial spreading. ICPhS 219-222 (1999).
[10] Schroeder, M.; Auberge, V.; Marie-Agnęs, C.: Can we hear smiles? Proc. 5th Int. Conf. Spoken Lang. Processing, Sydney, 559-562 (1998).
[11] Emond, C.; Trouvain, J.; Ménard, L.: Perception of smiled French speech by native vs. non-native listeners: A pilot study. Interdisciplinary Workshop on the Phonetics of Laughter, Aug. 4,5 (2007).
[12] Ruch, W.: Exhilaration and humor; in Lewis, Haviland, Hand book of Emotions, pp. 605-616 (Guilford Publications: New York, NY 1993)
[13] Lasarcyk, E.; Trouvain, J. Spread lips + raised larynx + higher F0 = Smiled Speech?- An articulatory synthesis approach. 8th International Seminar on Speech Production, pp. 345-248 (2008).
[14]Bickely, C.; Hunnicutt, S.: Acoustic analysis of laughter. Proc. Int. Conf. Spoken Lang. Processing, Banff, 927-930 (1992).
[15] Darwin, C.: The expression of the emotions in man and animals. (Oxford University Press (3rd Edition), Oxford 1998).
[16]De Benedictis, M.: Psychological and cross-cultural effects on laughter sound production. Interdisciplinary Workshop on the Phonetics of Laughter, Aug. 4,5 (2007).
[17] Titze, I.R.; Finnegan, E.M.; Laukkanen, A-M., Fuja, M., Hoffman, H.: Laryngeal muscle activity in giggle: A damped oscillation model. J. Voice. (2007).
[18]Esling, J. H.: States of the larynx in laughter. Interdisciplinary Workshop on the Phonetics of Laughter, Aug. 4,5 (2007).
[19] Estill, J.: Primer of Compulsory Figures (Estill Voice Training Systems: Santa Rosa, CA 1992).
[20] Dang, J; Honda, K.: Acoustic characteristics of the piriform ofssa in models and humans. J. Acoust. Soc. Am. 101:456-465 (1966).
[21] Imagawa, H.; Sakakibara, K. ; Tayama, N. ; Niimi, S.: The effect of the hypopharyngeal and supra-glottic shapes for the singing voice. Proc. Stockholm Musical Acoustics Conf. 2003 Vol. II: 471-474 (2003)
[22] Kitamura, T.; Honda, K.; Takemoto, H.: Individual variation of the hypopharyngeal cavities and its acoustic effects. Acoust. Sci. & Tech. (2004).
[23] Erickson, D.; Shochi, T..; Menezes, C.; Kawahara, H.; Sakakibara, K.: Some non-F0 cues to emotional speech: An experiment with morphing. Proc. Speech Prosody 2008 (2008).
[24] Erickson, D.; Yoshida, K.; Menezes, C.; Fujino, A.; Mochida, T.; Shibuya, Y.: Exploratory study of some acoustic and articulatory characteristics of *sad* speech. Phonetica 63: 1-25 (2006).
[25] Erickson, D.; Shochi, T.; Menezes, C.; Lu, X., Dang, J.: Some preliminary observations about articulatory, acoustic, and perceptual characteristics of *laugh* and *smile* speech in comparison with *sad* and *neutral* speech. Submitted to Mouton Special Edition.
[26] Kaburagi, T.; Honda, M.: Calibration methods of voltage-to-distance function for an electromagnetic articulometer (EMA) system *J.* acoust. Soc. Am. 111: 1414-1421 (1997).
[27] Honda, K.; Takemoto, H.; Kitamura, T.; Fujita, S.; Takano, S.: Exploring human speech production mechanisms by MRI. IEICE Trans. Inf. & Syst. E87-D: 1050-1058 (2004).
[28] Fujimura, O.: The C/D model and prosodic control of articulatory behavior. Phonetica 57:128-38 (2000).
[29] Erickson, D.: Articulation of extreme formant patterns for emphasized vowels. Phonetica 59: 134-149 (2002).
[30] Ruch, W.; Ekman, P.: The expressive pattern of laughter, In Kaszniak, Emotin, Qualia and Consciousness, 423-443 (World Scientifid, Tokyo,2001).

Ut. #	Intended emotion	phrase/word	%IDH	%IDS	%IDN	emotional intensity
1	H (ls)	**don't** go east	93%/**95%**	5%/**4%**	0%/**0%**	4.3/**4.2**
2	H (ls)	you **know**	85%/**84%**	7%/**10%**	0%/**0%**	3.1/**3.5**
3	H (ls)	do such a **thing**	82%/**83%**	13%/**11%**	0%/**0%**	3.9/**3.8**
4	H (ls)	gotta video tape **going**	91%/**81%**	6%/**16%**	0%/**0%**	3.5/**3.7**
5	H (ss)	**Oh** happy day	89%/**48%**	5%/**21%**	0%/**9%**	3.4/**2.9**
6	H	That makes me **sad**	8%/**17%**	40%/**41%**	32%/**26%**	1.9/**2.2**
7	H (ss)	mom, youre not paying **attention**	9%/**18%**	53%/**48%**	10%/**14%**	2.5/**2.4**
8	H (ss)	You **ask** my husband	5%/**18%**	69%/**57%**	3%/**7%**	3.1/**3.1**
9	S	I'd be very, I'd be very **sad**	3%/**0%**	94%/**98%**	**0%**/0%	4.3/**3.5**
10	S	you can **ask** for help	2%/**3%**	95%/**88%**	0%/**3%**	4.4/**3.6**
11	S	Pay **attention** dont	4%/6%	44%/**76%**	18%/**4%**	2.1/**2.9**
12	S	you **think**	6%/**4%**	57%/**76%**	12%/**7%**	2.3/**2.9**
13	S	**So**, I would miss her very much	2%/**9%**	91%/**8%**	1%/**40%**	3.3/**2.9**
14	S	You **know** it	0%/**2%**	84%/**39%**	5%/**47%**	2.9/**1.9**
15	S	like when somebody's **going** to leave you	0%/**1%**	97%/**19%**	0%/**64%**	4.4/**1.6**
16	S	Pay attention **dont**	4%/**8%**	44%/**13%**	18%/**52%**	2.1/**1.6**
17	N	think we're not **going** to get any real emoti	3%/**8%**	4%/**7%**	61%/**56%**	1.5/**1.6**
18	N	And I should have paid **attention** to the…	0%/**1%**	27%/**17%**	40%/**54%**	1.8/**1.6**
19	N	If I **ask** you questions	1%/**5%**	2%/**8%**	45%/**50%**	2.0/**1.7**
20	N	You **know**, I flew out on the 12th	3%/**3%**	4%/**13%**	70%/**49%**	1.4/**1.7**
21	N	happy and **sad**	0%/**0%**	7%/**42%**	50%/**44%**	1.7/**1.8**
22	N	And **so**, something that makes me feel sad	0%/**9%**	42%/**8%**	41%/**40%**	1.7/**1.8**
23	N	I **think** so too	3%/**4%**	36%/**27%**	37%/**38%**	1.9/**1.9**

Table 1. Results of perception tests for each phrase and each word. Column 1 shows the utterance number; column 2, the emotion intended by the speaker; column 3 the phrase and the word (in **bold**) used in the perception tests; columns 4, 5, and 6, the percent identification by listeners of the phrase or word (in **bold**) as either happy (%IDH), sad (%IDS) or neutral (%IDN) respectively to answer to question 2, and column 7 shows the listeners mean rating of the intensity of the emotion (answer to question 1) for the phrase and the word (in **bold**) where "5" indicates extremely emotional.