Some Articulatory Measurements of Real Sadness

Donna Erickson¹, Caroline Menezes² & Akinori Fujino³

¹Department of Intercultural Studies, Gifu City Women’s College, Gifu, Japan
² Laboratoire Parole et Langage, CNRS – Universite de Provence, Aix-en-Provence, France
³ NTT Communication Science Labs, NTT Corp, Kyoto Japan

erickson@gifu-cwc.ac.jp caroline.menezes@lpl.univ-aix.fr a.fujino@cslab.kecl.ntt.co.jp

Abstract

This study examines some of the articulatory differences in production of non-linguistic information, including spontaneous emotion vs. imitated emotion, para-linguistic information, in which the phrasing and intonational patterns were the same as in the first set of utterances with non-linguistic information, and linguistic information in which the same utterances were read. The findings suggest there are differences in articulation between imitated (acted) and spontaneous emotion. Specifically, for this particular emotional speech on the vowel /i/, tongue position was found to be higher and more forward, which suggests that emotional speech may be characterized by stronger lingual gestures.

1. Introduction

There is the general observation that what a speaker wishes to communicate is conveyed not only by linguistic units of syntactically concatenated phonemes and words, but by acoustic changes in the speaker’s voice including F0, duration, intensity, tone of voice, rhythm, phrasing, etc. As mentioned by Eldred and Price [1], communication involves not only “what” is said, but “how” it is said. Fujisaki [2] has suggested making a distinction between linguistic information (i.e., discrete categorical information which can be explicitly represented by the written language or uniquely inferred from context) from paralinguistic information (discrete and continuous information added by the speaker to modify or supplement the linguistic information) and non-linguistic information (information not generally controlled by the speaker, such as the speaker’s gender, age, etc.). This would also include “emotion,” both spontaneous and acted, even though with the later, the actor controls the expression of emotion. It has been shown (e.g., [3, 4]) that articulatory changes occur in acted productions of utterances in which the paralinguistic (sadness, or anger) and emotional (anger) content of the utterance varied. However, do acted emotions belong to a different category from those emotions being experienced “on-line” by the speaker? The actor controls the verbal expression in order to convey a certain affect to the listener, whereas the person who is experiencing the emotion while speaking probably does not have the same type of control. For instance, studies such as [5], showed hormonal endocrine (i.e., serum cortisol) changes as the result of sadness or elation. This suggests that the biological state of the speaker changes as a consequence of an emotion such as sadness or elation. It follows that laryngeal and supralaryngeal articulation may change also. A previous study by Erickson et al [6] suggested that acted and spontaneously expressed emotions are produced with different patterns of jaw and tongue articulations.

This study examines some of the articulatory differences in production of non-linguistic information, including spontaneous emotion vs. imitated emotion, para-linguistic information, in which the phrasing and intonational patterns were the same as in the first set of utterances with non-linguistic information, and linguistic information in which the same utterances were read. This study also asks the question whether there is a difference in articulation between “real” online spontaneous sadness and imitated sadness.

2. Methods

2.1. Data recording

As part of a larger study on emotion, both acoustic and articulatory data were recorded (2D EMA system, NTT Research Laboratories, Atsugi, Japan) during two sessions, five months apart. The subject was one American (Midwest dialect) female speaker. The first set of recordings was done as an informal spontaneous telephone dialogue with a friend through an earphone/microphone set-up, where the friend sat in a separate room from the subject. The friend asked the speaker various unrehearsed questions based on a list of topics related to the speaker’s personal life to evoke happiness, sadness or anger. The timing of the experiment was extremely fortunate for collecting sad (grieving) emotions (including crying while speaking); since the subject had just recently learned that her mother had been diagnosed with a fatal untreatable disease. The friend was very aware of this situation, and much of the interview focused on this particular situation in the subject’s life. The dialogue continued in a natural manner, while EMA recordings were made within a window frame of 20 sec, with a break in recording between frames. Acoustic recording, however, was continuous.

The second set of recordings done five months later provided control data for the initial data recordings. During this session, the subject performed three tasks: (1) she imitated the phrasing, intonation and emotion of the original utterances while at the same time listening to a taped recording through headphones, and looking at a transcript which she had marked indicating phrasing and intonation patterns of the original utterances, (2) she imitated only the phrasing and intonation, also while listening to the original utterance and looking at the transcript, and (3) she read the original utterance from the transcript marked for the phrasing and intonation of the original recording. Before the second session, the subject rehearsed the various tasks.
This report examines the articulatory characteristics of the word “really,” as uttered five (or four) times in a single 20 sec utterance, under four utterance conditions: (E) spoken in the first experiment as the speaker was grieving for her mother; (IE) imitating the phrasing, intonation and emotion while listening to a taped recording of the original utterance; (II) imitating only the phrasing and intonation, also while listening to the original utterance; (R) reading the original utterance, keeping the phrasing and intonation the same as in the original recording.

2.2 Articulatory analysis

For the EMA data analysis, movements were measured for the receiver coils attached to the lower incisor (mandible) and the second and third of four receiver coils (T2 and T3) attached along the longitudinal sulcus of the speaker’s tongue. The positions of the transmitter coils determine the coordinate system [7] with the origin positioned slightly in front of and below the chin. For the EMA data, all values are positive, with increasingly positive y-values indicating increasingly raised jaw or tongue dorsum position, and increasingly positive x-values, increasingly retracted jaw or tongue position. Articulatory measurements were made at the lowest vertical position of the mandible receiver coil in the syllable, using a MATLAB-based analysis program.

Recordings of the original spontaneous emotion (E) and of the control utterances--Imitated Emotion (IE), Imitated Intonation (II) and read speech (R)--were made in two sessions 5 months apart. In order to compare data of the two sessions, care was taken to minimize the differences of the coordinates of the palate shapes during the two sessions were analyzed, and the 5-degree difference was used to transform the coordinates of the first experiment with that of the other (see Erickson et al, 2003).

Sample articulatory recordings are shown in Figures 1a, b, c, and d (arranged from top to bottom) for the utterance “really, really, really, really, really” spoken on the four conditions: (a) spontaneous emotion E, (b) imitated emotion IE, (c) imitated intonation II, and (d) read R. Articulatory measurements were made during the mid point of the vowel of the second syllable /i/, usually at the point where the jaw was at lowest position. The dip in the jaw opening corresponding to the center of the vowel is difficult to see in this reduced-in-size figure. The vertical lines indicated by arrows above mark the point at which measurements were made.

Figure 1: “really, really, really, really, really” spoken on 4 conditions: (E)motion, (I)mitated (E)motion, (I)mitated (I)ntonation, (R)ead, respectively, from left column bottom, to right column top to right column bottom. The top two panels in each section show the x-y tracings of the mandible, the next two panels, those of the part of the tongue in front of the dorsum (T2), the next two panels, those of the tongue dorsum (T3), and the bottom two panels, the acoustic signal and spectrogram.
2.3 Perceptual rating judgments

The 22 instances of “really” were extracted from the recorded utterances, and presented auditorily by Sennheiser headphones to 10 English speakers using a Macintosh computer and the PsyScope Software. The listeners were asked to rate each “really” according to the perceived degree of sadness on a 5-point scale, where 1 indicated least sad and 5, most sad. Each instance of “really” was presented 4 times in a randomized order for a total of 88 trials self-paced. For each trial, the word was presented twice, with a gap of 1 sec. A short practice test of 6 utterances preceded the test.

3. Results

3.1 General observations

Looking at Figure 1 which shows the articulatory patterns for the utterance “really, really, really, really, really” spoken in four conditions, we can see a difference in the patterns of jaw opening. For spontaneous emotion (top panel), there tend to be two jaw openings during the second syllable of the word, with the first of the two corresponding to the production of the onset /l/ and the second to the middle of the /i/ vowel. The jaw opening corresponding to the onset /l/ tends to be larger than that for the vowel nucleus. However, for the other three conditions, we see one large jaw opening that starts during the first syllable, and ends during the second syllable. This is seen especially for the read version of the utterance (bottom panel). This unique pattern of two jaw openings for the second syllable of the emotional utterances may be a characteristic of emotional expressions which needs to be researched further.

3.2 Rating judgments

The results from 10 adult English listeners (9 Americans, 1 British) are shown in Table 1 and Figure 2. No category was rated 100% of the time as extremely sad (“5”) or extremely not sad (“1”), but in general, the ratings of sadness correspond to those intended by the speaker.

Table 1: Results of perception tests. “5” indicates very sad, and “1” indicates not sad at all. The first column indicates the condition of the utterance (Emotion, Imitated Emotion, Imitated Intonation, and Read), and the top row, the rating assigned by the 10 listeners to each “really” according to the condition.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>4%</td>
<td>21%</td>
<td>19%</td>
<td>32%</td>
<td>24%</td>
</tr>
<tr>
<td>IE</td>
<td>14%</td>
<td>22%</td>
<td>20%</td>
<td>28%</td>
<td>16%</td>
</tr>
<tr>
<td>II</td>
<td>27%</td>
<td>36%</td>
<td>28%</td>
<td>8%</td>
<td>0%</td>
</tr>
<tr>
<td>R</td>
<td>49%</td>
<td>29%</td>
<td>20%</td>
<td>2%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Figure 2: Averaged rating by English listeners of 22 utterances of the word “really”. The y-axis indicates the perceived degree of sadness on a scale of 1-5 (most sad); the x axis, the utterance condition categories: (e)motion, (i)mitated (e)motion, (i)mitated (i)ntonation, (r)ead.

3.3 Articulatory results

The results of the articulatory measurements are shown in Figure 3a, b, c. Figure 3a shows the jaw position for the different conditions. The jaw position for the read utterances (R) cluster together, but there is not a striking difference in jaw position for the other conditions. Figures 3b and 3c, however, show a striking difference especially for T2 as well as for T3 between the spontaneous sad utterances (E) and the other conditions: the tongue is higher and further forward.

A pattern of a more high, more forward tongue dorsum during production of a word with an /i/ vowel spoken on the word “leave” while actually crying (as compared to production of the same word in the other conditions, IE, II, and R) was reported in a previous analysis of the same data base (Erickson et al, 2003). However, during crying for the word “leave” (the E condition), the jaw was found to be decidedly lower than during the other conditions. It was suggested that the high forward tongue position was necessary in order to produce a high vowel in the environment of an open jaw. In this data, the E utterance with the most forward jaw position (Fig. 3a) also has the lowest, most back tongue dorsum position (Fig. 3c), which reflects the expected articulatory underpinnings of /i/ production. However, in this data we see a clear difference in tongue positions for the E-utterances, even though the jaw positions for the E-utterances are not strikingly different from those of the other conditions. Stronger lingual gestures may be a characteristic of emotional speech. More data analysis is needed to confirm these observations.
4. Conclusions

Perception tests showed that listeners were able to rate degrees of perceived sadness of the “really” utterances, and that in general, the spontaneous emotional utterances (E) were perceived as most sad, the imitated emotional utterances (IE), as next sad, the imitated intonational utterances (II) as not very sad, and the read utterances (R) as least sad.

The articulatory measurements showed that in terms of jaw position, the read utterances (R) clustered together, whereas no clear clustering was seen for the E, IE, and II conditions. However, tongue movement for the emotional utterances (E) was clearly different from that of the other conditions. Specifically, for this particular emotional speech on the vowel /i/, the tongue position was higher and more forward, which suggests that emotional speech may be characterized by stronger lingual gestures.

The findings of differences in articulation between acted and spontaneous emotion as reported in [6] as well as in this study suggest that there may be a difference between acted and spontaneous emotion in terms articulation. These findings underline the necessity in future studies for distinguishing between acted emotion vs spontaneous emotion as sub-types of non-linguistic information.

5. Acknowledgements

I thank Jianwu Dang for the MATLAB program used to analyze the articulatory data.

6. References


Figure 3: Pellet positions during production of “really” on 4 conditions. (a) Top shows jaw x-y positions (mm), (b) middle, T2x-y positions (mm), (c) bottom, T3 x-y positions (mm). The x-y axes are displayed so that the leftmost bottom corner represents forward and downward position of the jaw and tongue, as if the speaker were facing to the leftside of the page.