FRIENDLY SPEECH AND HAPPY SPEECH – ARE THEY THE SAME?

Lucy Noble & Yi Xu

University College London, UK
lucy.noble12@yahoo.com; yi.xu@ucl.ac.uk

ABSTRACT

In this study we explore the acoustic cues of friendly and happy speech by modifying naturally produced neutral-emotion utterances along a set of hypothetical Bio-informational Dimensions (BIDs), and using them as stimuli in a perception test. Subjects listened to these stimuli and judged their degree of friendliness and happiness. Results show that both expressions involve cues along the BID dimensions of *size projection* and *dynamicity*, but happiness is far more extreme along both dimensions. The significance of these results in relation to recent findings about the role of smile in social interactions is discussed.

Keywords: friendliness, happiness, bio-informational dimensions (BIDs), size projection, dynamicity

1. INTRODUCTION

Is happy speech the same as friendly speech? The answer may seem to be yes given the findings of two studies on the natural occurrence of smiles [7, 9]. It is demonstrated that, despite the general belief and previous evidence that the smile, especially Duchenne smile, [1] signals genuine happiness [6], people do not automatically smile when they are genuinely happy, but they smile much more often in social interactions. This has been interpreted as indication that the smile is mainly serving a social function rather than to express true happiness. These findings may have implications for emotional expressions in speech as well. It has been shown that listeners can correctly judge from speech alone whether the person speaking is smiling or what kind of smile it is, whether the smile is spontaneous [1, 5] or "mechanical" (i.e., without underlying emotions) [13], and that smiled speech is heard as happier [13]. The finding that smiling is social might suggest that happy speech is simply heard as being friendly. Alternatively, it is possible that listeners can still hear a difference between speech that is intended to be merely friendly and speech that sounds genuinely happy.

This study will investigate whether a difference can be heard between the two similar affective states, i.e., happiness and friendliness.

1.1. The acoustics of emotional expressions

Recent research has shown that an effective approach to understanding the acoustics of vocal expression of emotion is through the size projection principle, which was first proposed by Morton [10] as motivation-structural rules and further elaborated by Ohala [11] as frequency code. According to the principle, animals of many species signal aggressiveness by exaggerating their body size both visually and vocally. Visually they erect the hair or feather, standing erect or spreading out the wings. Some animals even develop permanent size markers such as the mane of lion and hump of bison. Vocally they emit calls with low pitch, rough quality [10] and lengthened vocal tract [11]. Also following the size projection principle, animals signal submission and appeasement by minimizing their body size, which not only indicate non-threat, but also elicit parental protection instinct by imitating infants. They retract the hair or feather, or crouching down [11]. Vocally they emit calls with high pitch, mellow voice quality [10] and shortened vocal tract [11]. In fact, it has been suggested that the human smile, which is likely homologous to the fear grimace of many primate species, is to shorten the vocal tract to project a small body size [11].

The body size projection principle has recently been found to be effective in predicting the perceptual contrast between anger and happiness. It is shown that vowels generated with an articulatory synthesizer with a lengthened or shortened vocal tract and lowered or raised F_0 are heard as sounding angry or happy [3]. It is further shown that the same perceptual effects can be achieved by modifying the density of the entire spectrum to simulate a lengthened or shortened vocal tract [14].

Effective as they are in the two-way forced choice identification tasks used in [3, 14], the stimuli do not sound markedly angry or happy.

This suggests that there may be other acoustic dimensions also involved in the encoding of the emotions. It is proposed in [14] that size projection is actually just one of a set of bio-informational dimensions (BIDs), evolutionarily developed under the selection pressure of interacting with other individuals, either conspecific or cross-species, that serve to *elicit behaviours that may benefit the vocalizer*:

Size projection — to project a large or small body size to create an effect of dominating or appeasing the receiver, so as to express threat/assertiveness, or friendliness/subordination. The encoding parameters are spectral density due to vocal tract length, F_0 and voice quality.

Dynamicity — controls the vitality of the vocalization, depending on whether it is beneficial for the vocalizer to appear strong or weak. A vigorous vocalization has a large movement range, in terms of both F_0 and formant movements, whereas a less vigorous vocalization has a narrow movement range.

Audibility — determines how far a vocalization can be transmitted from the vocalizer, depending on whether and how much it is beneficial for the vocalizer to be heard over long distance. The control of audibility is mainly through intensity. But it may also affect voice quality.

Association — controls associative use of sounds typically accompanying a non-emotional biological function in circumstances beyond the original ones. For example, the disgust vocalization seems to mirror the sounds made when a person orally rejects unpleasant food [4]. Articulating this kind of sounds involves tightening the pharynx, which would result in raised F1 as well as devoicing.

The advantage of the BID hypothesis is that it allows construction of testable predictions about specific emotions, and it also enables connection of findings that otherwise seem unrelated. The present study is designed to test whether systematic manipulations along two of the BIDs — size projection and dynamicity, can lead to the perception of friendly and happy speech, and whether there are clear differences between the parameters that are perceived to be the most appropriate for the two types of affective speech.

2. METHOD

The basic design is to use, as the base, English sentences spoken with different voice qualities by a native speaker in a neutral emotion, and then modify their overall pitch height, pitch range and spectral density through resynthesis to create the perception stimuli. English-speaking listeners were then asked to rate the stimuli in terms of friendliness and happiness. The advantage of using real speech as the base is that all the other acoustic signals not directly manipulated are kept as natural as possible. Voice quality is a parameter that has not been tested before as part of the size projection dimension, although it is originally proposed by Morton as one of the two major cues for animal calls [10]. Due to lack of effective technology to synthetically modify voice quality, however, we decided to use humanly produced voice quality types. It was predicted that friendliness and happiness both involve parameter values that project a small body size. Whether and how much their values agree would then tell us whether there is a difference between the two types of speech. The dynamicity manipulation is to test whether it is possible that friendly and happy speech differ only in terms of how vigorous they sound to the listeners.

2.1. Speech Materials

A male native English speaker in his twenties was recorded saying the sentence "We were away a year ago" in a neutral emotion, in three voice qualities: modal, breathy and tense. He was also asked to place focus on the word "away".

2.2. Preparation of stimuli

The three recorded utterances were manipulated in terms of three parameters, shown in the first three columns of Table 1. The manipulations were made with a specially written Praat [2] script that applied the "Change gender" function in Praat. The total number of stimuli were 4 (formant ratio) x 4 (pitch median) x 4 (pitch range) x 3 (voice quality) = 192.

Table 1: Parameter manipulations.

Formant Shift Ratio	Pitch Median (log scaled)	Pitch range factor (log scaled)	Voice quality
1.2	300	3	Modal
1.1	238.11	1.65	Breathy
1	188.99	0.91	Tense
0.9	150	0.5	

2.3. Procedure

Twenty-four native English speakers, 15 female and 9 male, with adequate hearing, took part in the experiment individually. Their age ranged from 21 to 63. The experiment took place in various quiet

locations using a laptop computer that ran an ExperimentMFC module in Praat.

The listeners heard the stimuli through headphones and were asked, in task 1, to rate how friendly each sentence sounded on a scale of 1-5, and in task 2, how happy each sentence sounded, also on a 1-5 scale. The stimuli were presented in random order, and listeners were unaware that they were hearing the same samples again. Half of the participants did task 1 first and the other task 2 first. They were allowed to listen to each sentence for up to three times.

3. RESULTS

The perceptual scores were analyzed with two repeated measures ANOVAs, with Voice quality, Formant shift ratio, Pitch median and Pitch range as independent variables. All the main effects are highly significant, which will be discussed in conjunction with the figures. Most of the two-way interactions are also significant, but we will not discuss them in this paper due to space limit.

Figs. 1-3 displays mean perceptual scores for friendliness and happiness as a function of voice quality, formant shift ratio, pitch median and pitch range. From all the graphs we can see that listeners are highly sensitive to voice quality (friendliness: $F(2,46) = 20.67$, $p < 0.0001$; happiness: $F(2,46) = 78.09$, $p < 0.0001$), and breathy voice is heard as both the most friendly and most happy. Somewhat surprisingly, the modal voice is heard as the least friendly and least happy, while tense voice is intermediate. Upon rechecking the stimuli, we noticed that the speaker who produced the base sentences did not use truly tense voice as we had wanted, and it actually contained some breathiness.

In regard to Fig. 1, the main effect of Formant shift ratio is significant for both friendliness ($F(3,69) = 4.20$, $p < 0.01$) and happiness ($F(3,69) = 4.50$, $p < 0.01$). Interestingly, however, for friendliness the 1.0 ratio, i.e., the original vocal tract length, has the highest mean score (3.21), whereas for happiness, the shortest vocal tract (1.2) has the highest mean score (3.35). Furthermore, for each formant shift ratio the divergence due to voice quality is larger for happiness, hence greater perceptual sensitivity, than for friendliness.

Fig. 2 shows that perceived friendliness and happiness both increase as pitch median increases (friendliness: $F(3,69) = 10.14$, $p < 0.0001$; happiness: $F(3,69) = 56.14$, $p < 0.0001$), but the rate of increase is much greater for happiness. This again shows greater perceptual sensitivity to happiness than to friendliness.

Figure 1: Mean perceptual scores as a function of *Voice quality* and *Formant shift ratio*.

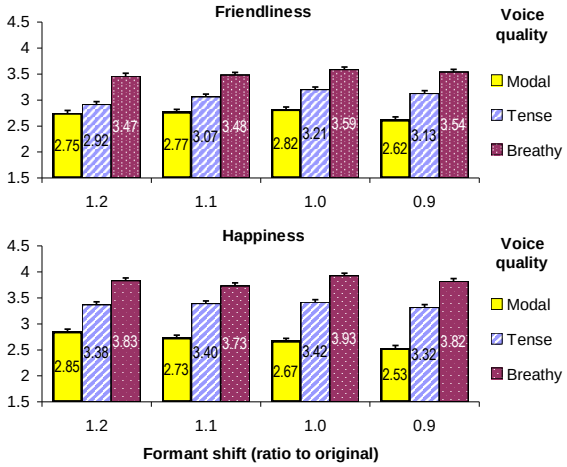


Figure 2: Mean perceptual scores as a function of *Voice quality* and *Pitch median*.

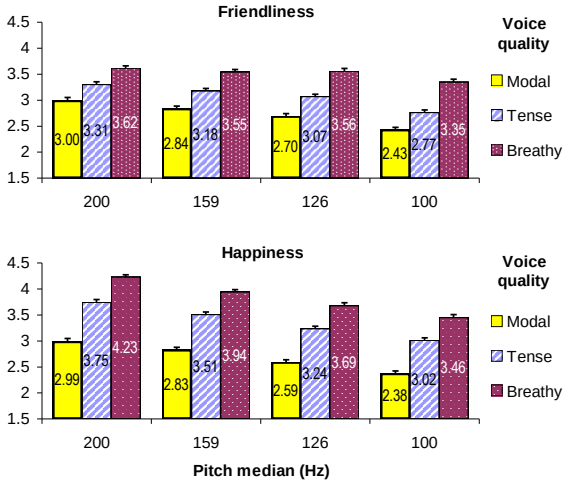
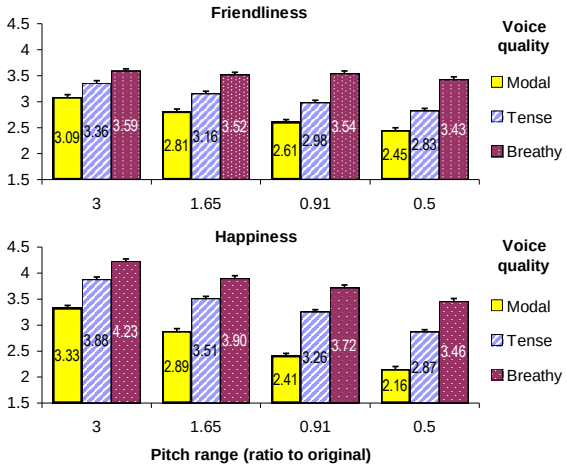


Figure 3: Mean perceptual scores as a function of *Voice quality* and *Pitch range*.



In Fig. 3, perceived friendliness and happiness again both increase with increased pitch range (friendliness: $F(3,69) = 5.96$, $p < 0.01$; happiness: $F(3,69) = 83.49$, $p < 0.0001$), but again the amount of increase is much larger for happiness than for friendliness, showing greater perceptual sensitivity to happiness than to friendliness.

Finally, the overall difference between friendliness and happiness was confirmed by a Wilcoxon signed ranks test, which showed that the difference was significant ($z = -6.055$, $p < 0.001$).

4. DISCUSSION

The present results show that listeners are highly sensitive to acoustic manipulations along two of the hypothetical BID dimensions — *size projection* and *dynamicity* (with the audibility and association dimensions left unmanipulated). The direction of their responses are also consistent with what would be predicted based on the meanings of the dimensions. Stimuli that project a small body size (breathy voice, high median pitch and shorter vocal tract) are heard as both happy and friendly, with the only exception that the original, rather than shortened, vocal tract was heard as the most friendly. Happiness perception was found to be sensitive to *dynamicity*, as predicted by the BID hypothesis, which is also consistent with various previous findings [12].

Despite their similarities, clear differences were also seen between friendliness and happiness. Listeners seem to expect happiness to be much more extreme along the *size projection* dimension. And, interestingly, they seem to expect friendly speech to involve normal rather than shortened vocal tract length. An important implication of this finding is that it could be the case that the previously reported social function of the smile [7, 9] is actually to generate a sign of happiness in the eye of the viewers, which is presumably much more engaging than simply being friendly. In contrast, friendliness is probably more akin to politeness, for which showing too much happiness may not be appropriate, at least in the cultural environment in which the current study is situated.

Perhaps the most surprising finding of the study is listeners' very high sensitivity to voice quality for both friendliness and happiness. This is despite our difficulty in getting the source speaker to produce the right tense voice, as explained earlier. The finding is consistent with [8], but here the effects seem much more robust. One possibility of such high sensitivity is due to the fact that voice quality is not used as a major cue for any linguistic contrast in English. But this needs to be investigated in future research.

Finally, the findings of this study demonstrate the effectiveness of using the BID hypothesis to investigate emotional expressions in a systematic manner, which makes it possible to establish mechanistic links between the findings of different studies.

5. REFERENCES

[1] Auberge, V., Cathiard, M., 2003. Can we hear the prosody of smile. *Speech Communication* 40, 87-97.

[2] Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341-345.

[3] Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., Maneewongvatana, S. 2008. Encoding emotions in speech with the size code. *Phonetica*. 65, 210-230.

[4] Darwin, C., 1872. *The Expression of the Emotions in Man and Animals*. London, England: John Murray.

[5] Drahota, A., Costall, A., Reddy, V. 2008. The vocal communication of different kinds of smile. *Speech Communication* 50, 278-287.

[6] Ekman, P., Davidson, R. J., Friesen, W. V. 1990. The Duchenne smile: emotional expression and brain physiology II. *Journal of Personality and Social Psychology* 58, 342-353.

[7] Fernandez-Dols, J.-M., Ruiz-Belda, M.-A. 1995. Are smiles a sign of happiness?: Gold medal winners at the olympic games. *Journal of Personality & Social Psychology* 69, 1113-1119.

[8] Gobl, C., Chasaide, A. N. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40, 189-212.

[9] Kraut, R.E., Johnston, R. E. 1979. Social and emotional messages of smiling: an ethological approach. *Journal of Personality and Social Psychology* 37(9), 1539-1553.

[10] Morton, E. S. 1977. On the occurence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist* 111(981), 855-869.

[11] Ohala, J. J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, 1-16.

[12] Scherer, K. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 227-256

[13] Tartter, V. C., Braun, D. 1994. Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America* 96, 2101-2107.

[14] Xu, Y., Kelly, A., Smillie, C. (forthcoming). Emotional expressions as communicative signals. In Hancil, S., Hirst, D (eds.), *Prosody and Iconicity*.

[1] Duchenne smile involves not only the retraction of the corners of the lips but the forming of wrinkles around the eyes [7].