PRODUCTION AND PERCEPTION OF SMILING VOICE

ILARIA TORRE

University of York

Abstract

This study investigates phonetic aspects in the production and perception of Smiling Voice, i.e. speech accompanied by smiling. A new corpus of spontaneous conversations is recorded to compare the formant frequencies of Smiling Voice (SV) and Non-Smiling Voice (NSV); the hypothesis that smiling raises formant frequencies is proven to be valid also on spontaneous speech, after previous research found this hypothesis to be true for scripted speech. Then, two perception experiments are carried out to test the hypothesis that listeners can recognize instances of SV extracted from the corpus of spontaneous data and instances of NSV obtained from read speech. Once the hypothesis is confirmed, a second perception experiment is performed to attempt to locate the point, in an artificial continuum from SV to NSV, where such perception happens.

1. Introduction

1.1. Smiling

The co-occurrence of speech and smile has been given various names in the literature, such as "speech-smile" (Kohler 2007), "smiled speech" (Émond and Laforest 2013) or "Smiling Voice" (Pickering et al. 2009). Here, the name Smiling Voice will be adopted, as opposed to Non-Smiling Voice (which describes neutral speech produced with no accompanying facial gesture); the abbreviations SV and NSV will sometimes be used.

Smiling is a universal phenomenon, common in humans and other animals (Mehu and Dunbar 2007: 271; van Hooff 1975: 231), which involves many movements in the facial area, especially in the region of the eyes, brows and mouth (Mehu and Dunbar 2007: 270). Ultimately, smiling shortens the vocal tract (Shor, 1978: 89), which affects the properties of the sound produced simultaneously.

1.2. Smiles and emotions

Smiles are mostly considered expressions of positive emotions such as happiness or joy (Kohler 2007: 21), although it is possible to fake smiles (Ekman and Friesen 1982: 244). Whatever emotion smiles can convey, humans seem to be quite good at recognizing it. This has been demonstrated with a number of experiments using only audio stimuli (Laukka 2005), only visual stimuli (Eisenbarth and Alpers 2011), non-verbal audio stimuli (Hawk et al. 2012) or crossed audio-visual stimuli (de Gelder and Vroomen 2000), in an application of the McGurk Effect, i.e. a phenomenon caused by a mismatched audio-video set of stimuli, which leads to the perception of an alien sound (McGurk and MacDonald 1976).

The present study focuses on some phonetic aspects of the production and perception of Smiling Voice. Even though the question of how Smiling Voice is linked to emotions was not addressed in any phase of the research, it is not excluded that emotions might have played a role in some of the experiments, especially the perception ones; after all, most of the study concentrates on naturally-occurring conversations, where emotions are put on display constantly.

Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2013**) conference

ISSN 1758-0315 Issue 1

2. *Production of Smiling Voice*

2.1. *Previous findings on the acoustics of Smiling Voice*

Some of the research on Smiling Voice includes the work by Kohler (2007: 21), who found that smiling increases the formant frequencies of the sound. Tartter (1980) found that smiling increases the formant frequencies, fundamental frequencies and, at least for some speakers, duration and amplitude, of the speech portion. These findings were confirmed by Tartter and Braun (1994), Drahota et al. (2008) and Fagel (2009). However, works on Smiling Voice are still scarce, and their methodology has yet to be proven valid. For example, most of the literature on the topic has used scripted speech, i.e. instructing actors or naïve speakers to artificially produce Smiling and Non-Smiling Voice. One of the few works employing natural conversations (Drahota et al. 2008) analysed the effect of smiling on vowels, without taking into account whole utterances, but it is clear that the acoustics of smiling affect whole portions of speech. Only Émond and Laforest (2013), in their very recent work on the prosodic correlates of Smiling Voice, used whole smiled words coming from a corpus of spontaneous conversations, but – since it was impossible for them to find a non-smiled counterpart for each word in that corpus – they chose other non-smiled words that had similar characteristics (e.g., same number of syllables and same duration). It is evident, however, that it would be pointless to make a comparison of the formant frequencies between such words, because the segments constituting the original words will be different.

The present research explores the hypothesis that changes in formant frequency from neutral voice to Smiling Voice affect data coming from naturally-occurring conversations as well.

2.2. *Methodology*

Five pairs of friends (four females and six males), all native English speakers, participated in two recording sessions. They were initially audio and video recorded while same-sex pairs conversed in a sound-proof booth of the University of York, using two Sanyo Xacti video recorders, one Zoom H4n audio recorder and two Beyer Opus 55 headset Microphones, at a sampling frequency of 44.1 kHz. In this first recording session, participants could choose to play a number of games that were designed to elicit smiles (e.g. picking objects from a box and recall experiences related to them (as used by Beskow et al. 2009: 190), playing "Taboo" (a card game that involves describing words), completing children's crossword puzzles, and playing rhymes).

Picture 1: The recording booth[1]

[1] Written permission to use audio-visual material from the recordings was obtained by all participants prior to the recordings, in accordance to the guidelines specified by the Ethics Committee at the University of York.

After the session, all the videos were watched in order to identify the utterances that contained smiles all throughout. Only single words or short intonational phrases (identified as chunks of speech associated with one intonation pattern, as described in Wells 2006: 6) were selected. Words or phrases which overlapped with the other participant's speech, contained instances of laughing voice or laughter, or were accompanied by background noise were discarded. When selecting the utterances that contained Smiling Voice, a smile was defined as a facial gesture involving a rising of the cheek muscles, a widening of the mouth and an upwards curving of the lips, as described by Shor (1978: 88).

The selected words and intonational phrases were then isolated in the corresponding audio files and extracted using the software Praat (Boersma and Weenink 2012), thus obtaining one short audio file for each instance of Smiling Voice.

The same subjects were called shortly afterwards to participate in a second recording session, where they were asked to read the list of words or phrases that they had originally uttered with Smiling Voice. They were instructed to read in a manner that was "as neutral as possible", without smiling. In this way, a significant number of pairs of SV and NSV were obtained, and the mean frequencies of the first three formants of each token were computed running a script in the software Praat. Even though it is clear that these pairs differed in more elements than just smiling (amplitude and f0, to name a few), since they were uttered in completely different contexts, the focus of this research was on the differences among the first three formants in the target words. Therefore, the reading modality was chosen as one of the ways to obtain exactly the same word that had originally been uttered in the spontaneous conversations.

2.3. Results

The final data set was composed of 3462 frequency values (1154 SV/NSV tokens * 3 formant values). These were computed both in Hertz and in Bark scales (using the formula provided by Traünmuller 1990: 99), in order to investigate both the production and perception aspects of Smiling Voice. A 2-tailed t-test was performed on the data to obtain the mean difference between frequencies in SV and NSV, then for each formant, then only for tokens containing rounded segments, and then for gender-specific differences. A discussion of all the results is given in paragraph 2.4.

Figure 1 summarizes the results of the first tests. The top charts show that, in general, the formant frequencies in Smiling Voice are higher than in Non-Smiling Voice. The frequencies of the first formant were found to be generally higher in Smiling Voice, while Non-Smiling Voice resulted in higher second formant frequencies. Finally, the frequencies of the third formant turned out to be much higher in Smiling Voice, with a difference of 53 Hz and 0.111 Bark.

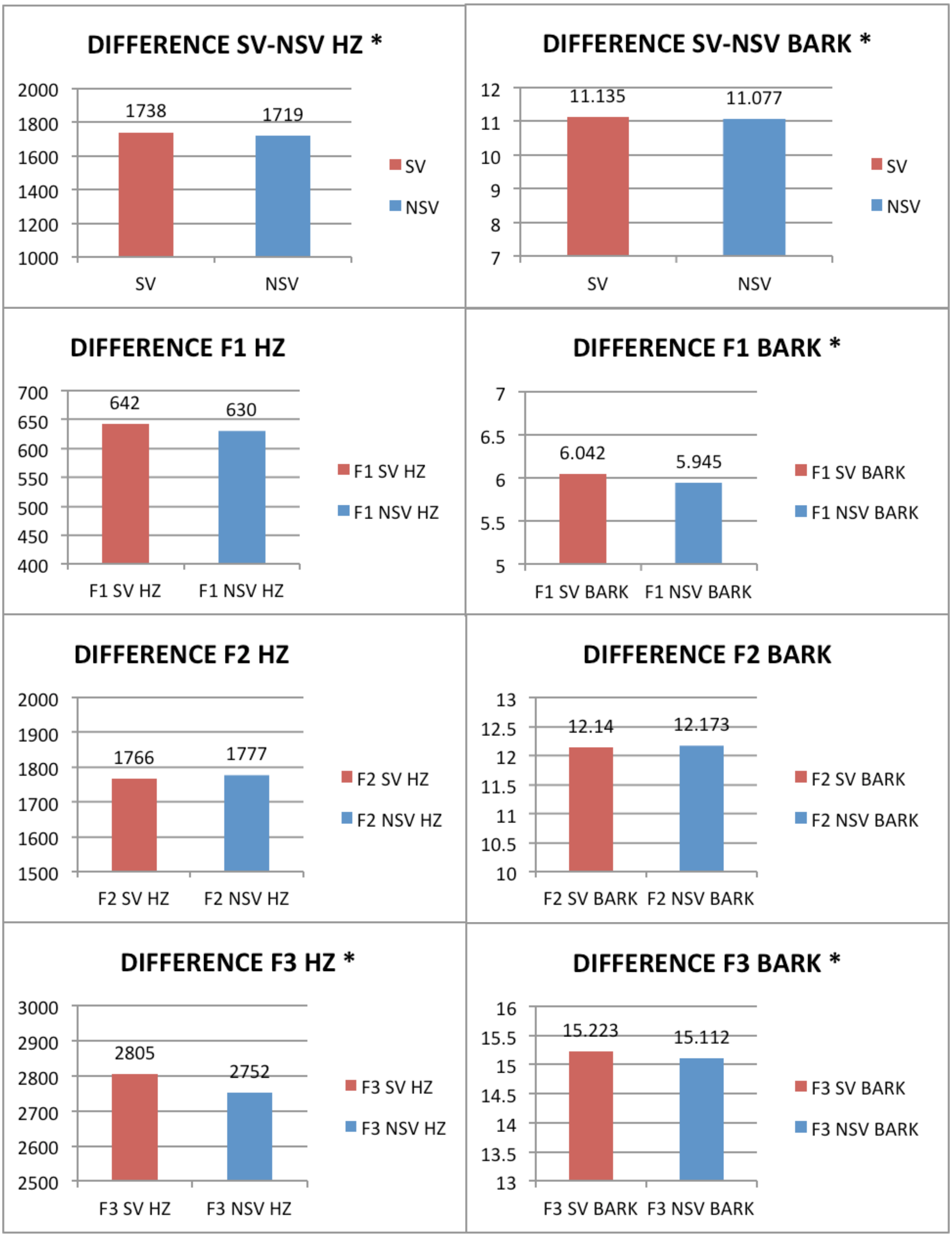

Figure 1: Differences in the average formant frequencies between Smiling and Non-Smiling Voice in Hertz (top left) and Bark (top right), and in f1, f2 and f3 frequencies.

Figure 2 represents the results of the test on the words that contained rounded vowels and/or consonants. The result was a mean difference of 73.7 Hz (top charts in figure 2.4 below), which is approximately seven times the mean difference calculated for all the words that did not contain rounding.

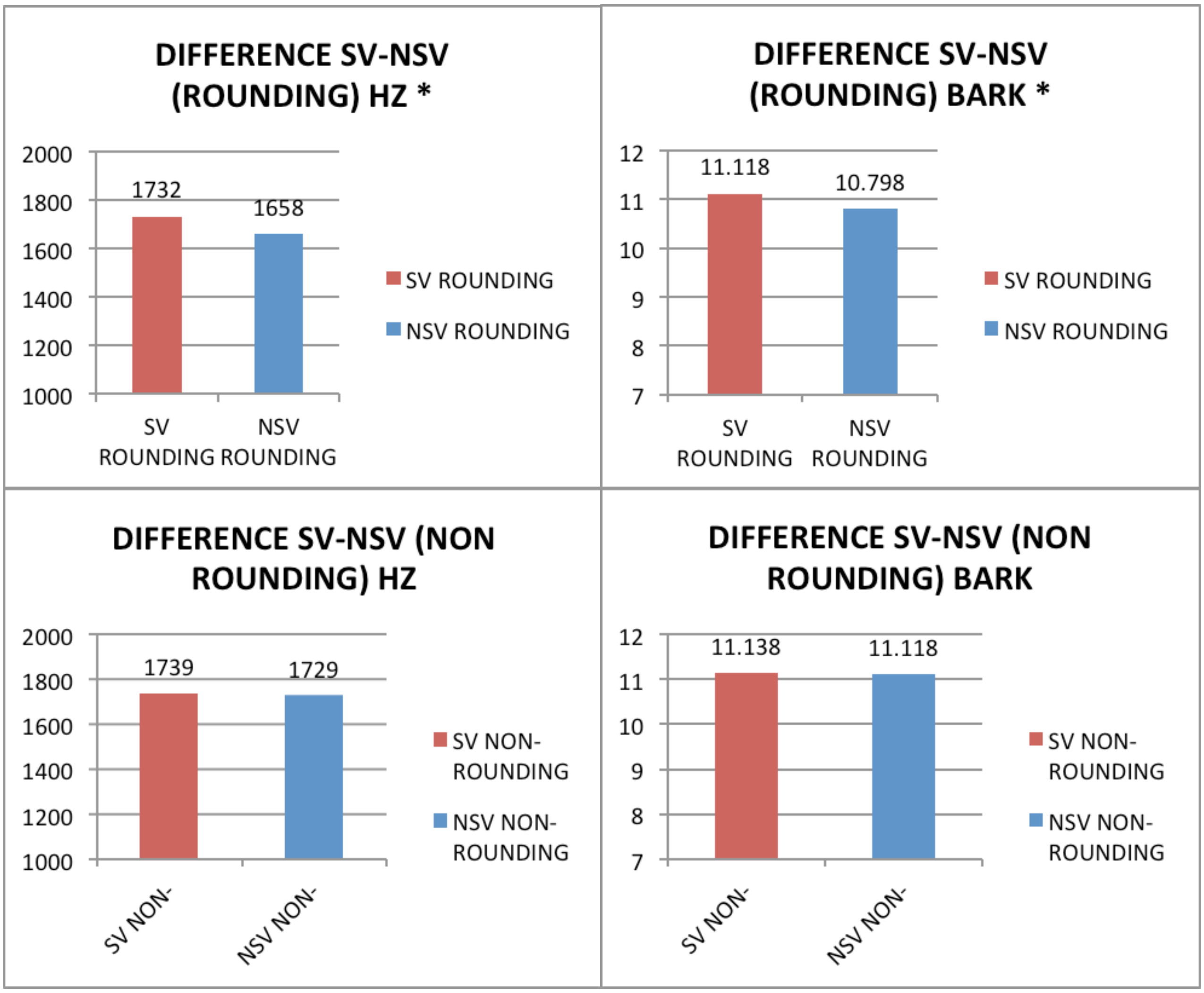

Figure 2: Differences in the average formant frequencies between Smiling and Non-Smiling Voice, in words containing rounding (top row) and words not containing rounding (bottom row).

Finally, Figure 3 shows differences in the data uttered by male speakers in comparison with the data collected from female speakers. Apart from showing that females' formant frequencies are higher than males', as expected (Clark et al. 2007: 269), it shows that, while the results for the male speakers are consistent with the average findings of Figure 1, female speakers present the reverse pattern, with frequencies in Non-Smiling Voice being higher than in Smiling Voice.

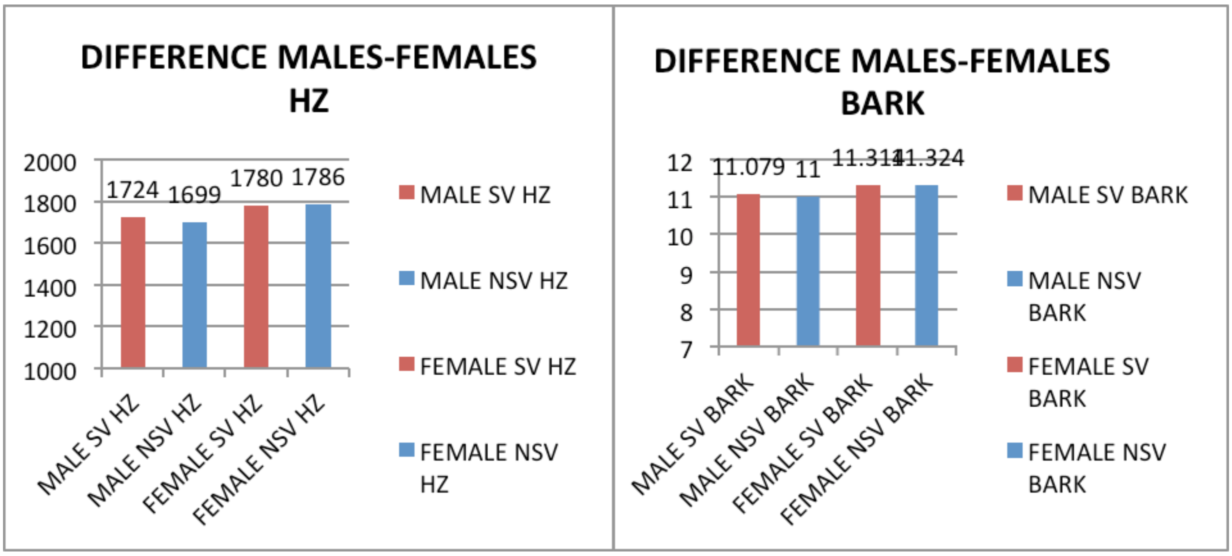

Figure 3: Differences in the average formant frequencies between Smiling and Non-Smiling Voice in male and female speakers.

2.4. Discussion

The general finding of this study is consistent with the hypothesis that smiling raises the formant frequencies of speech. This result appears to be statistically significant on the whole ($p<0.005$) and for words containing rounding ($p<0.005$). This second finding is not surprising, as lip rounding – which is usually accompanied by lip protrusion, and which lengthens the vocal tract (Fant 1960:116) – is the phenomenon that is affected the most by smiling – which, on the other hand, as said before shortens the vocal tract.

Examination of the individual formant frequencies, however, shows significant differences from previous experiments. In particular, Tartter (1980: 26) found that all three formant frequencies were higher in the smiling register, and considered f2 to be the most significant. Therefore, in a following study, Tartter and Braun (1994: 2105) reported only f2. In this study, however, f2 turned out to be the least significant value.

Differences in f1 appear not to be significant if the values are considered in Hertz ($p = 0.008$), but they start to be if the values are considered in Bark ($p = 0.005$), which means that, in general, differences in f1 in the passage from smiling to non-smiling are perceivable by a human ear. On the other hand, and contrary to previous findings, differences in f2 are negative, which means that, in general, Smiling Voice in this corpus has lower f2 frequencies than Non-Smiling Voice. However, these differences appear not to be statistically significant in both Hertz, and Bark ($p = 0.119$ and $p = 0.213$ respectively). What is striking is the comparison with Tartter (1980: 26), who found that, for 67% of his speakers, differences in f2 were statistically higher for smiling modality ($p < 0.002$), and with Lasarcyk and Trouvain (2008: 346), who found that lip spreading raises mainly the frequencies of f2. Since the second formant is the resonance associated with the length of the vocal tract, it would have been predictable that a shortening in the vocal tract would have raised the formant frequencies. However, in Tartter's research only scripted speech was used, which may have therefore biased the acoustics of the final data. Furthermore, neither the present study nor Tartter's employed a large enough number of speakers to make statistically relevant claims.

Instead, differences in f3 are statistically significant both in Hertz and in Bark ($p < 0.005$), and they represent the highest difference in the corpus (53 Hz and 0.111 Bark). This suggests that, of the three formants, f3 is the one that changes the most from smiling to non-smiling. This finding confirms previous results: Tartter (1980: 26) found that the difference from the scripted Smiling and Non-Smiling Voice was higher in f3 as well. However, she did not provide an explanation for it, and, in another study she chose to concentrate on f2, because the similar effect on the three formants was considered “redundant” (Tartter and Braun 1994: 2104).

Finally, there is a significant difference between the male and female speakers used in the recordings for this study. Even though, as expected (Huber et al. 1999: 1540; Clark et al. 2007: 269), females’ formant frequencies are generally higher than males’, what is striking is the relative difference between SV and NSV. Males have generally higher formant frequencies for Smiling Voice than Non-Smiling Voice, whereas females have generally lower formant frequencies for Smiling Voice than Non-Smiling Voice. This result might suggest that the male speakers employed in the present recordings tend to smile “more broadly” than the female speakers, but no such tendency was observed in the videos. Also, as only a small number of speakers were used in this study, it would be unreliable to account for this claim. Furthermore, a broad smile is not produced only by a change in the vocal tract (and formant frequencies changes refer exclusively to changes in the vocal tract), but also by using many different facial muscles, therefore requiring different amounts of energy for individual speakers. Such gestures, as already explained, are not considered here.

3. Perception of Smiling Voice

3.1. Previous research

Although, as mentioned above, smiling does not necessarily correspond to a particular emotion, it is true that some emotions can be expressed by smiling, and most of the studies carried out so far have concentrated on the perception of smiles as linked to an emotion. In particular, it has been found that factors such as a listener’s age (Lambrecht et al. 2012: 535; Paulmann et al. 2008: 265) or gender (Van Strien and Van Beek 2000: 650; Paulmann et al. 2008: 267) can influence said perception.

In this research, importance has been given to gender differences in the perception of Smiling Voice, which has been studied in two experiments. In the first one, subjects listened to a set of stimuli and were asked to recognize which stimuli were uttered while smiling and which ones were not. The second experiment constitutes an attempt to find the moment - in an artificial continuum from Smiling Voice to Non-Smiling Voice - where smile perception starts or ends.

3.2. Experiment 1

As mentioned above, the first perception experiment of this research seeks to confirm the hypothesis that listeners can recognize words uttered in Smiling Voice coming from spontaneous conversation, as opposed to words that were read without smiling.

3.2.1. Methodology

The stimuli for this experiment were taken from the recordings done initially to study the acoustics of Smiling Voice (section 2.2). They were selected from the words and phrases that had been already extracted from the first recordings, in pairs (i.e. for each word in Smiling Voice there was a corresponding word in Non-Smiling Voice). The final set was composed of

41 stimuli, 19 of which were filler items. 16 listeners, 8 males and 8 females, all native English speakers, volunteered to do the experiment, which took place in a quiet room, one person at a time. The audio was played on a laptop, and the listeners wore a noise-reducing Sennheiser HD 280 Pro set of headphones. They were given an answer sheet with 41 boxes, and were asked to fill each box with an S (if they thought that the corresponding word had been uttered with a smile) or with an N (if they thought that the corresponding word had been uttered without smiling). The listeners' response times were not recorded, but none of them asked to listen to a stimulus twice, and all of them managed to write each answer in the time slot provided (i.e. the 6 second pause between one stimulus and the next). No participant withdrew from the experiment, and no boxes on any answer sheet were left blank.

3.2.2. Results and discussion

As Figure 4 shows, the number of correct answers surpasses by far the number of incorrect ones: the participants answered correctly 77.9% of the time. This figure shows that listeners actively recognized Smiling and Non-Smiling Voice, without simply guessing (a normal distribution test confirmed that their performance was significantly better than chance, $z > 1.645$).

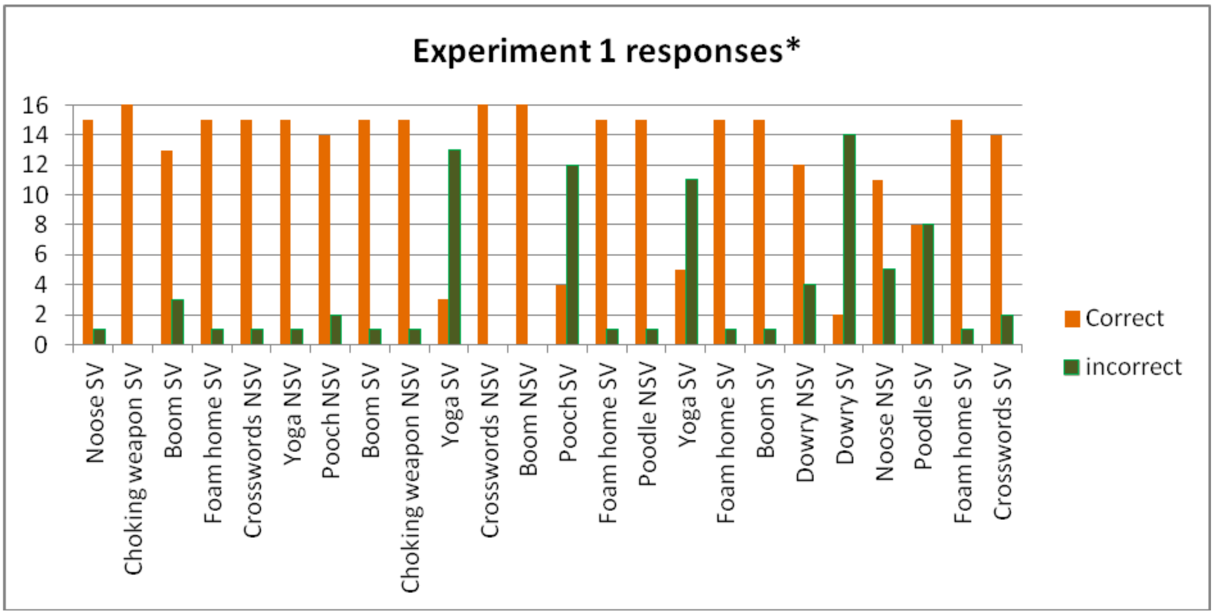

Figure 4: Responses to all the stimuli of experiment 1

It is possible to see that four stimuli in particular seemed to be difficult to classify for the participants. These are *Yoga* (SV), *Pooch* (SV), and *Dowry* (SV). These three words have in common the fact that they were uttered by a male, and that they present overall different prosodic features from the rest of the target words. In particular, all of them are marked by a low-falling pitch contour, and one of them (*Dowry*) is uttered with creaky voice. The speaker's gender can be excluded from being an influence on the listeners' perception, because other smiled target words uttered by males were recognized as such by the majority of listeners. What seems to change in these four words is that their prosodic characteristics resemble some of the features of read speech, which is what the non-smiling data set is made of. Figure 5 shows the pitch contours of the smiled and neutral versions of the three problematic words, in comparison with the smiled and neutral versions of three other target words that seemed not to cause any problems in the listeners (shown in Figure 6).

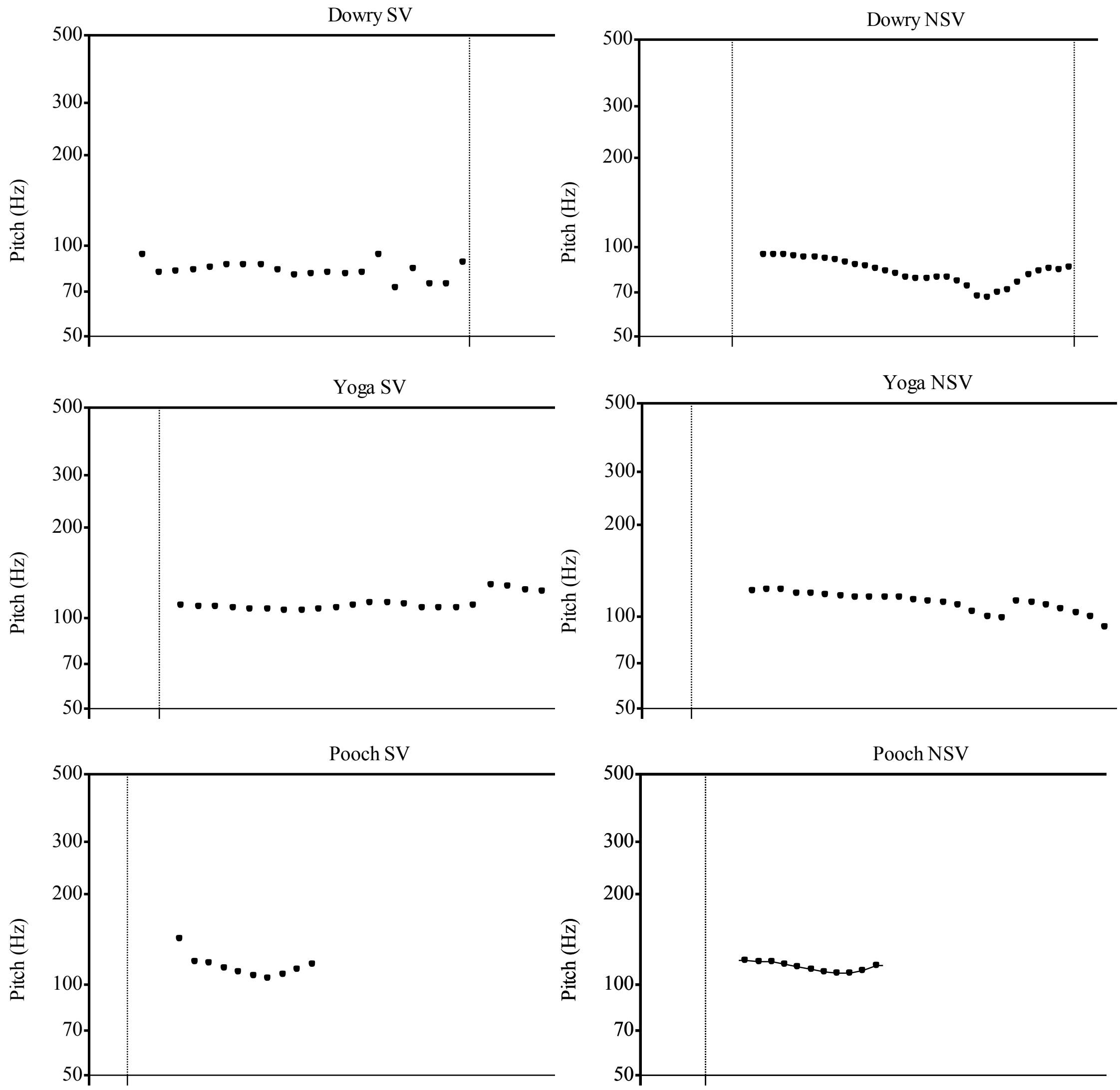

Figure 5: Pitch contours of three problematic words in experiment 1.

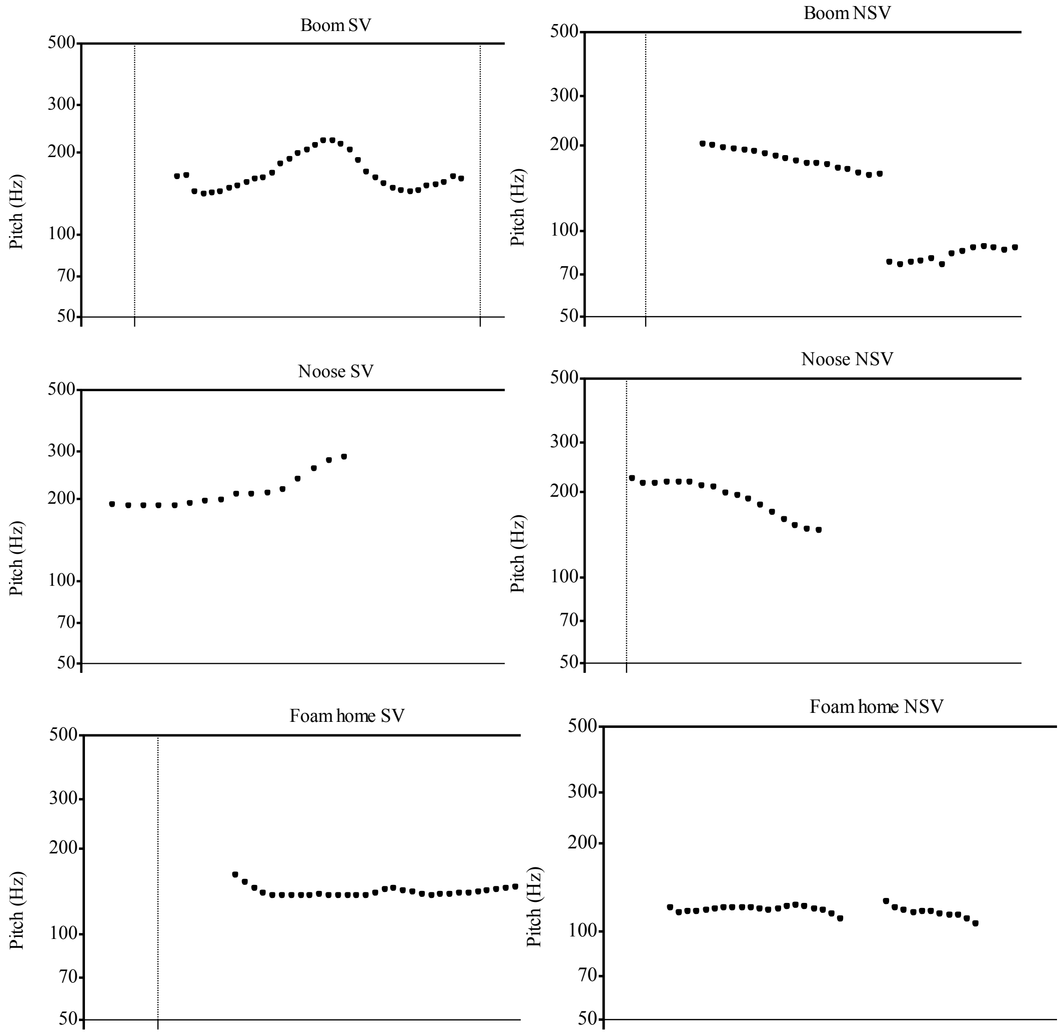

Figure 6: Pitch contours of three unproblematic words in experiment 1.

There is a growing amount of literature on the prosodic differences between spontaneous and read speech. For example, Laan and van Bergem (1993: 572) found that prosody plays a vital role in the differentiation between read and spontaneous speech: in their experiment, they artificially manipulated the f0 frequencies of read and unscripted utterances, so that one contained the frequency of the other, and found that listeners' performances significantly diminished. Similarly, the listeners' answers in the present experiment might have been biased by the fact that all the other words that they had classified as N presented similar prosodic characteristics, whereas the words that they classified as S tended to show a higher amount of variation in their suprasegmental features. This variation, according to Vaissière (2005: 252), seems to be associated with emotions such as happiness and pleasantness, which, in turn, tend to be associated with smiling. A low, monotonous fundamental frequency and slower speaking rate, on the other hand, tend to be associated with sadness and boredom (Vaissière 2005: 252; Lasarcyk and Trouvain 2008: 345). These results, however, partly contrast with Aubergé and Cathiard (2003: 95). In their data, what changed the most from spontaneous to (in their case) acted speech was amplitude, rather than f0. Their spontaneous data, however, was strictly controlled, and some of their speakers were professional actors. Also, their data was entirely in French, and there might be some language-specific differences in the use of intensity in the carrying of spoken emotions. Instead, Lasarcyk and Trouvain (2008: 347) found that the main cue in the recognition of lip spreading on single vowels was f0; but they were aware that using single vowels would not provide complete results, and pointed out that future research should use longer utterances. They did not exclude the possibility that their listeners associated each stimulus with an emotion, but neither in their research nor in the present one were emotions mentioned to listeners in the perception experiments. They suggested that future research should use smaller f0 manipulations in the perception experiments, and, in a way, that is what has been done in the present research: f0 has not been manipulated at all, but it was originally different in the SV-NSV pairs.

All these considerations, however, are made on a small number of occurrences, and should be only intended as an attempt to explain the anomalies in the particular perception experiment carried out here.

Apart from these particular cases, the listeners' responses (summarized in Figure 7) show that they were better, on average, at recognizing Non-Smiling Voice than Smiling Voice. An independent samples t-test showed that this result is statistically significant (p=0.000).

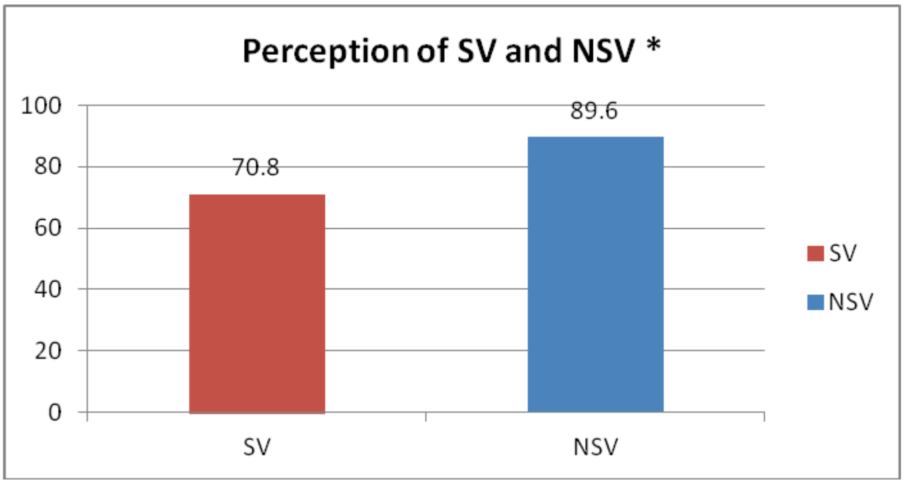

Figure 7: averages of correct identification of Smiling and Non-Smiling Voice.

As for possible gender differences, Figure 8 shows that females were generally better than males at recognizing both Smiling and Non-Smiling Voice. An independent sample t-test for equality of means, however, showed that these differences were not statistically significant (p > 0.1): there is no evidence that one gender performed significantly better than the other.

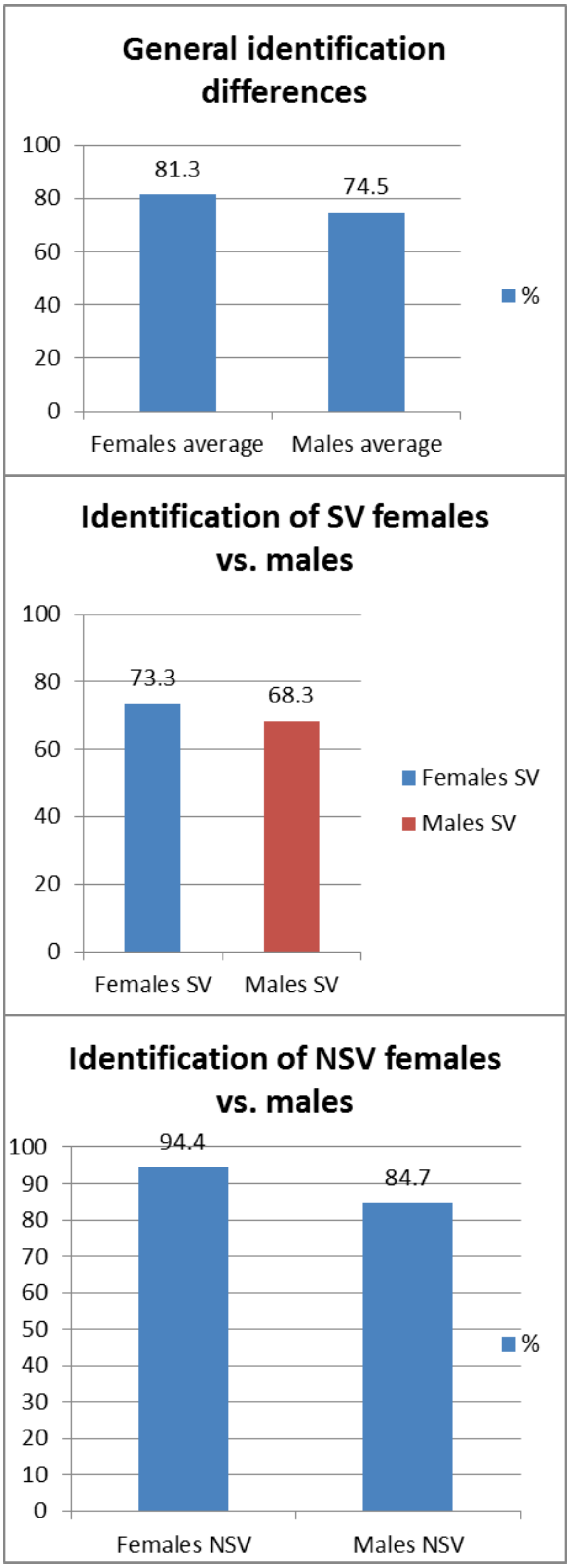

Figure 8: Gender differences in recognizing SV and NSV.

These results seem to be consistent with the findings of some previous studies on gender differences in perception. For example, Schirmer et al. (2002) found that female listeners performed better than males in perception experiments, even though in a second study (Schirmer et al 2005; Paulmann et al. 2008: 267) they found that gender disparities are eliminated if listeners are told to take into account prosody when doing the task. Similarly, an experiment involving nonsense words whose different prosodic characteristics carried different emotions (Pell et al. 2009: 430) proved that prosody is an essential cue for distinguishing emotions in speech.

In the present experiment, participants were only instructed to make a distinction between Smiling Voice and Non-Smiling Voice, and it is therefore possible that one of the cues that they were listening to was prosody. This would confirm the impression that the anomalies in the four words that were mistaken by the majority of listeners were due to a specific prosodic pattern. Listeners were asked to recognize smiles, and it is possible that they associated smiles with happiness, leaving the purely acoustic characteristics of the speech that they were hearing in the background.

3.3. Experiment 2

The second perception experiment, once confirmed that listeners can recognize Smiling and Non-Smiling Voice, seeks to determine at which point, in a continuum from smiling to neutral modality, such recognition happens.

3.3.1. Methodology

A native English speaker (female, aged 24) was audio recorded while reading six words. She was instructed to read each word twice, once with a straight face and once with stretched lips, but keeping the same intonation and, as far as possible, the same duration. This was necessary because of the configuration of Akustyk[2] , the programme that was used to synthetize a 5-step sound continuum from the "smiled" word to the neutral word. After many trials, it became evident that Akustyk could not create sound continua of files containing obstruent or voiceless sounds, or if the original starting and ending words had different intonation contours or very different durations. For these reasons, the words to be used had to contain voiced, non-obstruent sounds, and a similar prosodic pattern. The words that were chosen are: *Arrow, Loan, Mole* and *Wool*. Two more words (*Pool* and *Spoon*) were used as fillers. For this experiment, only words containing rounding were used because, since the formant frequencies show a wider variation from smiling to non-smiling modality in rounded words than in unrounded ones, it would have been easier for Akustyk to create intermediate steps that were as far as possible from each other.

Figure 9 shows the variations in the trajectories of the first three formants in the continuum of one of the target words. The three horizontal lines represent the formant trajectories, and the vertical columns of red dots represent the steps that Akustyk created from the formant frequencies of the start word to the formant frequency of the end word.

[2] Akustyk is an extension for Praat that allows to synthetize speech in a number of different ways, including creating a speech continuum. It was downloaded from http://bartus.org/akustyk in April 2013. The software is no longer available to download from the author's website, but it is free and open source, and it is possible to copy a version of it from the computer of someone who already has it.

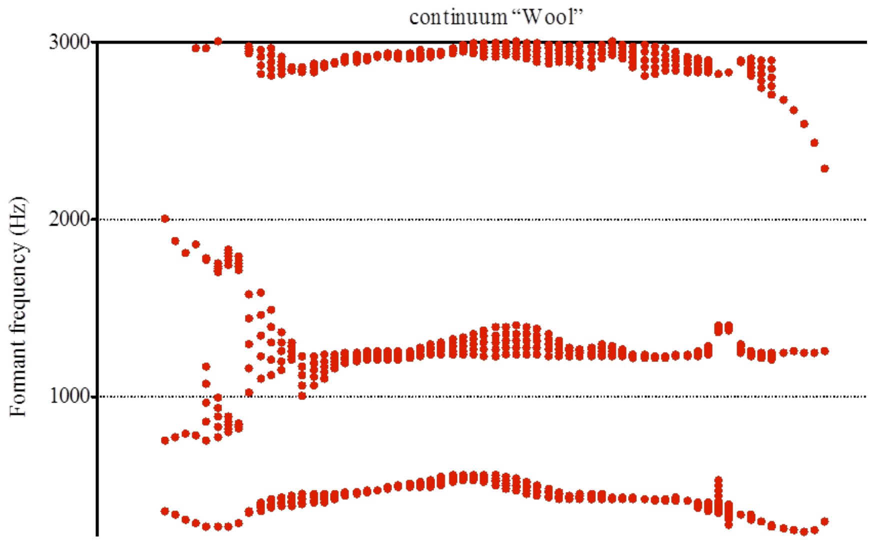

Figure 9: Continuum SV-NSV for the word W*ool*

16 native English speakers (8 males and 8 females), different from the participants in experiment 1, volunteered to take part in the experiment, which followed the same procedure as experiment 1.

3.3.2. Results and discussion

As it is possible to see from Figure 10, the experiment responses were far from uniform. In three out of four cases, the majority of listeners could at least identify the first step in the continuum as being smiled; but, for the case of *Mole*, only 3 listeners out of 16 (18.75%) correctly identified the first step as being smiled. As for the rest, the general tendency was that the majority of stimuli were identified as Non-Smiling. During the de-briefing, many subjects reported to the researcher that they had found it very difficult to distinguish any difference at all.

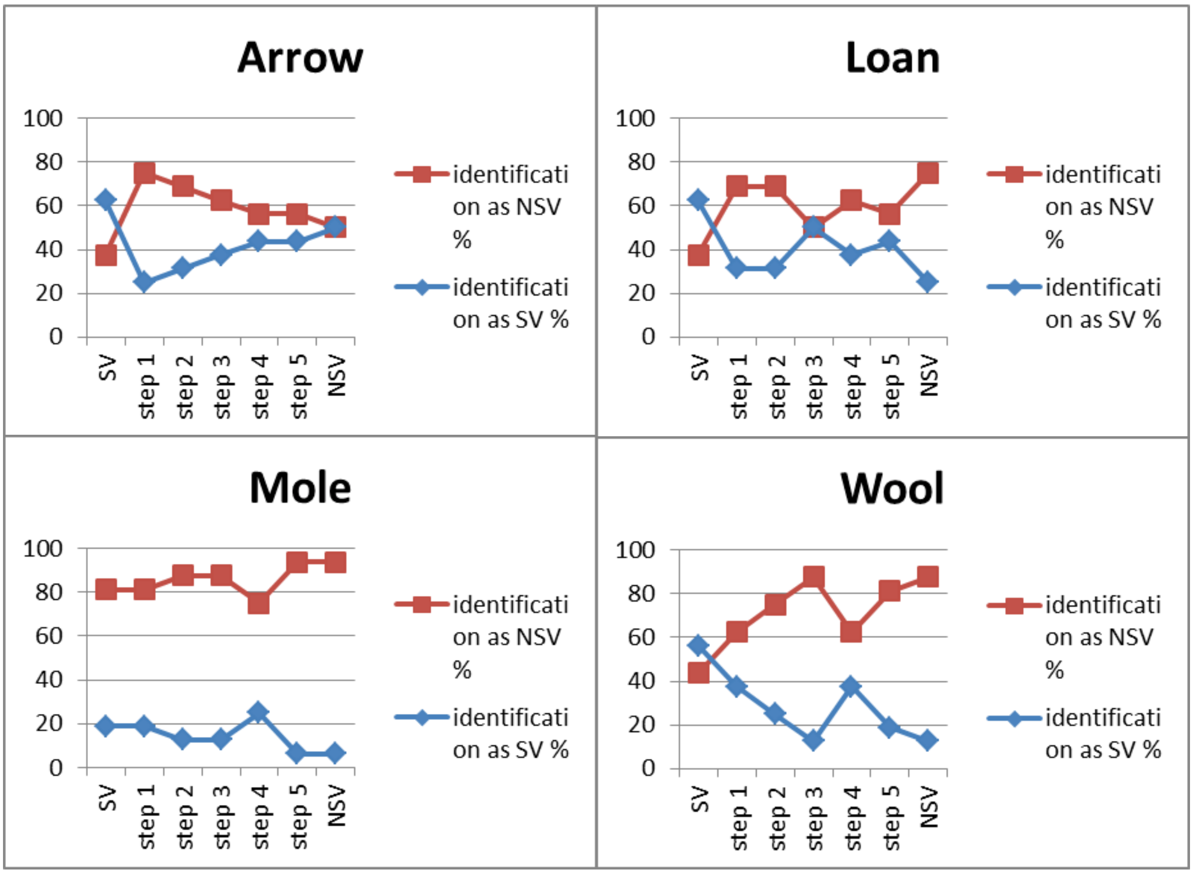

Figure 10: Trend answers for each of the target words of experiment 2.

These results could be due to the fact that the intermediate steps, which were artificially produced, were not realistic enough, or to the fact that the quality of the Akustyk's outputs was still too poor. Although the quality of the manipulation might be a good explanation for the difficulty of this experiment, it is also possible that listeners relied on prosodic cues to operate a distinction between SV and NSV. Since Akustyk was unable to create a continuum from two words with different intonations, in the final outputs only the formant frequencies were changed. Therefore, the intonation remained the same, since the speaker who was recorded uttered all her words with the same, low-falling intonation.

However, the stimuli used in this experiment were too few, and the quality of the synthesis was too poor, to allow to draw such a conclusion. More experiments are needed, with better synthesizer tools and more stimuli, before drawing a general conclusion. The present results contrast with Robson and MacKenzieBeck (1999), who found that listeners could recognize the effects of lip spreading on speech, and associate it with smiling. In their experiment, however, the stimuli were presented to the listener in an orderly succession, and they had to choose between pairs of stimuli, so that they were progressively trained in recognizing the differences. In the present experiment, instead, the stimuli order was randomized, and there were no cases of close steps being presented one after the other.

3.4. General discussion

Experiment 1 confirms that speakers can recognize Smiling Voice without being presented with visual stimuli, and supports the hypothesis that this recognition happens even when the audio stimuli comes from unscripted speech.

Experiment 2 does not present any significant result on the location of the recognition of Smiling Voice in an artificial continuum from Smiling Voice to Non-Smiling Voice. However, another interesting question is posed: is prosody essential in the perception of Smiling Voice? Future research will have to find an answer.

4. Conclusion

The present study examined some phonetic characteristics of the phenomenon of Smiling Voice.

It was found that, as predicted by the model of the vocal tract and by past literature, instances of Smiling Voice taken from spontaneous speech increase the frequencies of the first three formants, and that this increase is statistically significant. Differences were found among the individual formants: the most noticeable difference turned out to be in f3, and the least noticeable difference turned out to be in f2. Differences were also found in relation to the gender of the speaker, with the difference in the formant frequencies between Smiling and Non-Smiling Voice being higher in male speakers than in females.

To investigate the auditory correlations of Smiling Voice, two original perception experiments were carried out. Experiment 1 confirmed that listeners can distinguish words uttered in smiled and read modalities without many difficulties, and that, if difficulties arise, they are probably due to a lack of prosodic cues to differentiate the two modalities. The importance of prosody is stressed again in Experiment 2, where the responses indicated that the majority of stimuli were identified as non-smiled, even when they had originally been uttered with a smiling gesture. This suggests that prosody plays a vital role in the recognition of Smiling Voice, and that more accurate synthesizers (or experiments involving a manual synthesis of speech) should be used to investigate the matter further.

It is possible that the importance of prosody is due to the fact that prosody contributes to carrying emotional meaning in speech, and that listeners tend to associate smiling with an emotion. For the same reason, future research should also make sure that listeners do not make use of semantic cues for the recognition of Smiling Voice, e.g. by linking a particular stimulus to an emotion. This problem could be solved by using stimuli in smiling and neutral modalities with different degrees of prosodic variation coming from languages that the listener does not know. Another path to follow that the present research did not pursue was analysing data coming exclusively from spontaneous speech. In fact, although in this study it was possible to use spontaneous instances of Smiling Voice, the corresponding instances of Non-Smiling Voice were not spontaneous, as the speakers were reading from a list.

References

AUBERGÉ, V. AND CATHIARD, M. 2003. Can we hear the prosody of a smile? *Speech Communication 40*, pp. 87–97.

BESKOW, J., EDLUND, J., ELENIUS, K., HELLMER, K., HOUSE, D. AND STRÖMBERGSSON, S. 2009. Project presentation: Spontal - multimodal database of spontaneous dialog. *Proceedings of Fonetik 2009, the 12th Swedish Phonetics Conference. University of Stockholm, Sweden, 10-12 June 2009, Department of Linguistics, Stockholm University,* pp. 190-193.

BOERSMA, P. AND WEENINK, D. 2012. Praat: doing phonetics by computer [on line]. Version 5.3.53, available from http://www.praat.org/ [October 2012].

CLARK, J., YALLOP, C. AND FLETCHER, J. 2007. *An introduction to Phonetics and Phonology. Third edition.* Oxford: Blackwell.

DE GELDER, B. AND VROOMEN, J. 2000. The perception of emotions by ear and by eye. *Cognition and emotion 14 (3)*, pp. 289-311.

DRAHOTA, A., COSTAL, A. AND REDDY, V. 2008. The vocal communication of different kinds of smile. *Speech Communication 50*, pp. 278–287.

EISENBARTH, E. AND ALPERS, G. W. 2011. Happy Mouth and Sad Eyes: Scanning Emotional Facial Expressions. *Emotion 11 (4)*, pp. 860-865.

EKMAN, P. AND FRIESEN, W. V. 1982. Felt, False, and Miserable Smiles. *Journal of Nonverbal Behavior 6 (4)*, pp. 238-252.

ÉMOND, C. AND LAFOREST, M. 2013. Prosodic correlates of smiled speech. *Proceedings of Meetings on Acoustics, ICA 2013 Montreal, Montreal, Canada, 2 - 7 June 2013.*

FAGEL, S. 2009. Effects of Smiled Speech on Lips, Larynx and Acoustics. *Proceedings of AVSP 2009, Norwich, Sept 10th-13th*, pp. 18-21.

FANT, G. 1960. *Acoustic Theory of Speech Production*. Mouton, The Hague.

HAWK, S. T., FISCHER, A. H. AND VAN KLEEF, G. A. 2012. Face the Noise: Embodied Responses to Nonverbal Vocalizations of Discrete Emotions. *Journal of Personality and Social Psychology 102 (4)*, pp. 796-814.

HUBER, J. E., STATHOPOULOS, E. T., CURIONE, G. M., ASH, T. A. AND JOHNSON, K. 1999. Formants of children, women, and men: The effects of vocal intensity variation. *Journal of the Acoustical Society of America 106 (3)*, pp. 1532-1542.

KOHLER, K. J. 2007. ‘Speech-smile’, ‘speech-laugh’, ‘laughter’ and their sequencing in dialogic interaction. *Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter, Saarbrücken, 4-5 August 2007*, pp. 21-26.

LAAN, G.P.M. AND VAN BERGEM, D. R. 1993. The contribution of pitch contour, phoneme durations and spectral features to the character of spontaneous and read aloud speech. *Proceedings of Eurospeech '93, Berlin*, pp. 569-572.

LAMBRECHT, L., KREIFELTS, B. AND WILDGRUBER, D. 2012. Age-related decrease in recognition of emotional facial and prosodic expressions. *Emotion 12 (3),* pp. 529-539.

LASARCYK, E. AND TROUVAIN, J. 2008. Spread Lips + Raised Larynx + Higher F0 = Smiled Speech? - An Articulatory Synthesis Approach. *Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France, 8-12 December 2008*, pp. 345-348.

LAUKKA, P. 2005. Categorical Perception of Vocal Emotion Expressions. *Emotion 5 (3),* pp. 277-295.

MCGURK, H. AND MACDONALD, J. 1976. Hearing lips and seeing voices. *Nature 264*, pp. 746-748.

MEHU, M. AND DUNBAR, R. I. M. 2007. Relationship between Smiling and Laughter in Humans (Homo sapiens): Testing the Power Asymmetry Hypothesis. *Folia Primatol 79*, pp. 269-280.

PAULMANN, S., PELL, M. D. AND KOTZ, S. A. 2008. How aging affects the recognition of emotional speech. *Brain and Language 104*, pp. 262-269.

PELL, M. D., PAULMANN, S., DARA, C., ALASSERI, A. AND KOTZ, S. A. 2009. Factors in the recognition of vocally expressed emotions: a comparison of four languages. *Journal of Phonetics 37*, pp. 417-435.

ROBSON, J. AND MACKENZIEBECK, J. 1999. Hearing smiles - perceptual, acoustic and production aspects of labial spreading. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, 1-7 August 1999*, pp. 219-222.

SCHIRMER, A., KOTZ, S. A. AND FRIEDERICI, A. D. 2002. Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research 14*, pp. 228-233.

SCHIRMER, A., KOTZ, S. A. AND FRIEDERICI, A. D. 2005. On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research 24*, pp. 442-452.

SHOR, R. E. 1978. The production and judgment of smile magnitude. *Journal of General Psychology 98*, pp. 79-96.

TARTTER, V. C. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics 27 (1),* pp. 24-27.

TARTTER, V. C. AND BRAUN, D. 1994. Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America 96 (4),* pp. 2101-2107.

TRAÜNMULLER, H. 1990. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America 88 (1),* pp. 97-100.

VAISSIÈRE, J. 2005. Perception of Intonation. In PISONI, D. B. AND REMEZ, R. E. (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell Publishing, pp. 236-263.

VAN HOOFF, J. A. R. A. M. 1975. A Comparative Approach to the Phylogeny of Laughter and Smiling. In HINDE, R. A. (ed.) *Non-Verbal Communication*. Cambridge: University Press, pp. 209-241.

VAN STRIEN, J. W. AND VAN BEEK, S. 2000. Ratings of Emotion in Laterally Presented Faces: Sex and Handedness Effects. *Brain and Cognition 44*, pp. 645–652.

WELLS J. C. 2006. *English intonation: an introduction*. Chapter 1, Cambridge: CUP, pp. 1-14.

Ilaria Torre
Department of Language and Linguistic Science
University of York
Heslington
York
Email: ilaria.torre11@gmail.com