PAN
ARCHIVES OF ACOUSTICS
Vol. **44**, No. 1, pp. 13–26 (2019)
DOI: 10.24425/aoa.2019.126348

Perceptual Identification of Polish Vowels Due to F0 Changes

Mariusz OWSIANNY[(1), (2)]

[(1)] *Institute of Linguistics*
Adam Mickiewicz University in Poznań
Al. Niepodległości 4, 61-874 Poznań, Poland; e-mail: marows@amu.edu.pl

[(2)] *Poznań Supercomputing and Networking Center*
Jana Pawła II 10, 61-139 Poznań, Poland; e-mail: mowsianny@man.poznan.pl

(*received May 8, 2018; accepted October 22, 2018*)

The paper investigates the interdependence between the perceptual identification of the vocalic quality of six isolated Polish vowels traditionally defined by the spectral envelope and the fundamental frequency F0. The stimuli used in the listening experiments were natural female and male voices, which were modified by changing the F0 values in the ±1 octave range. The results were then compared with the outcome of the experiments on fully synthetic voices. Despite the differences in the generation of the investigated stimuli and their technical quality, consistent results were obtained. They confirmed the findings that in the perceptual identification of vowels of key importance is not only the position of the formants on the F1 × F2 plane but also their relationship to F0, the connection between the formants and the harmonics and other factors. The paper presents, in quantitative terms, all possible kinds of perceptual shifts of Polish vowels from one phonetic category to another in the function of voice pitch. An additional perceptual experiment was also conducted to check a broader range of F0 changes and their impact on the identification of vowels in CVC (consonant, vowel, consonant) structures. A mismatch between the formants and the glottal tone value can lead to a change in phonetic category.

Keywords: F0; formants; speech perception; vowel shifts; voice quality.

1. Introduction

The vowel space as defined by the frequencies of the first two formants for contemporary Polish is sparsely filled. However, it has not always been this way. In the Middle Ages, Polish had additional narrowed vowels, i.e. /á/, /é/, /ó/ (the grapheme ó exists to this day). In the earlier Proto-Slavic period, the Proto-Polish vowel system included reduced vowels called yers (Długosz-Kurczabowa, Dubisz, 2006), which underwent significant shifts. In contemporary Polish, there are only six oral vowels, which are represented by the phonemes: /i/, /y/, /e/, /a/, /o/, and /u/. Despite such a small repertoire of vowel sounds, the mismatch between vowel formant frequencies and the glottal tone F0 may result in a change in phonetic category, which has been proven in a number of listening experiments conducted by the author (Owsianny, 2001). To generate test signals, a formant speech synthesizer has been used. Results of these numerous and laborious measurements of isolated synthetic vowels, published in PROSODY 2000 conference proceedings, have been included with the consent of the Polish Phonetic Association in the present paper to compare them with the current data for natural vowels with the manipulated fundamental frequency F0.

Fant's (1960) theory of speech production is based on the assumption that the source signal (i.e. the vibration of vocal folds) is independent from the bank of acoustic filters, i.e. the resonances of the vocal tract. However, the frequency of the glottal tone F0, which describes the work of vocal folds and the harmonics defined by this frequency, have an impact on both real and imaginary formants. Real formants result from the resonances of the vocal tract, while imaginary formants result from the integration of formants and particular harmonics in the centre of gravity, COG (Chistovich *et al.*, 1979). Naturally, due to their anatomical structure, female and child voices have higher formant values than male voices. Raised F0 in the phonation process and tensed vocal folds and larynx muscles lead to increased formant frequencies through the raising of the larynx (the thyroid cartilage) and, as a result, additional shortening of the vocal tract (Obrębowski, 2008). Any change in voice pitch involves a change in the formant frequencies F1, F2 (Chladkova *et al.*, 2009) that have an impact on the phonological and interpersonal variation between vowels (Jassem, 1992; Peterson, Barney, 1952). Chladkova *et al.* (2009) prove that a change in the formant frequencies due to F0 changes is significantly greater for female voices and apply more to the first formant F1. Intended F0 change causes changes in formant values both in short and long Czech vowels. However, the greatest formant change occurs in open vowels. This effect is unrelated to the aforementioned anatomical and physiological dependencies. The authors suggest that this phenomenon can only be attributed to the spectral undersampling hypothesis (Diehl *et al.*, 1996), which concerns the formant dependence on F0 and the F1 × F2 vowel plane expansion for female voices and the consequent compensatory effect of retrieving information that has been lost due to increased gaps between harmonics. They estimate the percentage rate for retrieved information caused by increased F1 values due to increased F0 values for open vowels, which is 64 percent. This hypothesis seems to work only for relatively small F0 value changes (for female voices, F0 was changed only within the 200–280 Hz range). It also appears that the possibility of changing formant values adapted to F0 changes is specifically restricted by the size of the vocal tract and the limited range of articulator movements. Diehl *et al.* (1996) also emphasize additional, i.e. behavioural, reasons for a greater between-category dispersion of female vowels on the F1 × F2 plane. This is related to the sexual dimorphism hypothesis which is based on Darwin's argument that acoustic differences between male and female phonemes become more prominent as both genders want to demonstrate greater distinctiveness. The authors also attempt to find a reason for this nonuniform scaling and for weaker vowel identification due to an increased F0 frequency. On the other hand, what stands in some contrast to the effect of vowel space expansion for female voices with higher F0 values than for male voices is the interdependence of the mean acoustic distance (in dB) between English vowels and the fundamental frequency, which shows a notable decrease with increased F0 for both male and female voices (Diehl *et al.*, 1996). A lower mean acoustic distance with increased F0 results in weakened identification and worse perception for high F0 values, which is commonly observed. Assamann and Nearey (2008) claim that "the gradual decline in accuracy as a function of both upward and downward spectral envelope shifts and the interaction between spectral envelope shifts and F0 suggests the additional operation of perceptual mechanisms sensitive to the statistical covariation of F0 and formant frequencies in natural speech" (p. 3203). Thus, the previously mentioned hypotheses that attribute decreased identification values solely to an increased F0 are far from sufficient. This point will be confirmed by author's findings discussed in the latter part of this paper. Assamann and Nearey (2008) prove that when F0 and the spectral envelope scale factor are shifted in the same directions (they both grow or decline), vowel identification is more precise than in the cases when they are shifted in the opposite directions. Those studies, however, concern vowels in general, not individual phonemes, which will be the subject of inquiry in the present study.

The effect of increased formant values in relation to the glottal tone and harmonics can easily be observed in the female voice pronouncing a vowel /a/ in the spectrogram in the Fig. 1. How do the values of the first two formants change as a function of time with an increased F0? At first, they increase significantly and then the formants start to overlap with the lower harmonics. For high F0 values, it is very difficult to precisely determine the formant values. It is only feasible to attempt to estimate these values.

Besides the observed impact of the F0 frequency on the values of the lowest formants, another subject of inquiry was the impact of F0 on vowel perception. Perceptual studies by Fant (1960), Carlson *et al.* (1975), Chistovich *et al.* (1979), Di Benedeto (1994), Diehl *et al.* (1996), Imiołczyk (1991), Johnson (1988a; 1988b), Syrdal (1985), Traunmüller (1981), Hirahara and Kato (1992) and others prove a considerable impact of the fundamental frequency F0 on the perceptual vowel quality. It has been found that the same structure of the formants can cause them to be perceived as two different vowels depending on the frequency of the excitation source. Modifying the position of the formants, especially the lowest ones, F1 and F2, with the constant voice pitch, causes a change in vowel quality. Greater changes have a distinguishing impact on the phonetic category. By way of analogy, small manipulations of the fundamental frequency F0 change the sound of the vowel through the impact on the perception of its opening. Greater F0 changes (after exceeding a critical value) can result in a change in phonetic category, as has been proven by Carlson *et al.* (1975), Syrdal (1985) and Owsianny (1995, 2001). This effect, which is a function of vowel similarity in a given language, is conditional on the density of the vowel space. The formants of a given vowel need to be correlated with the appropriate fundamental frequencies. If their F0 is too low, the perceived vowel becomes more open. Conversely, if F0 is too high, the vowel is perceived to be more closed (Traunmüller, 1981). The effect of a perceptual vowel shift due to the glottal tone frequency changes occurs most frequently in female voices (Owsianny, 1995). They are much more susceptible to F0 changes and the number of misidentified vowels with considerable F0 changes is approximately three times higher than in male voices.

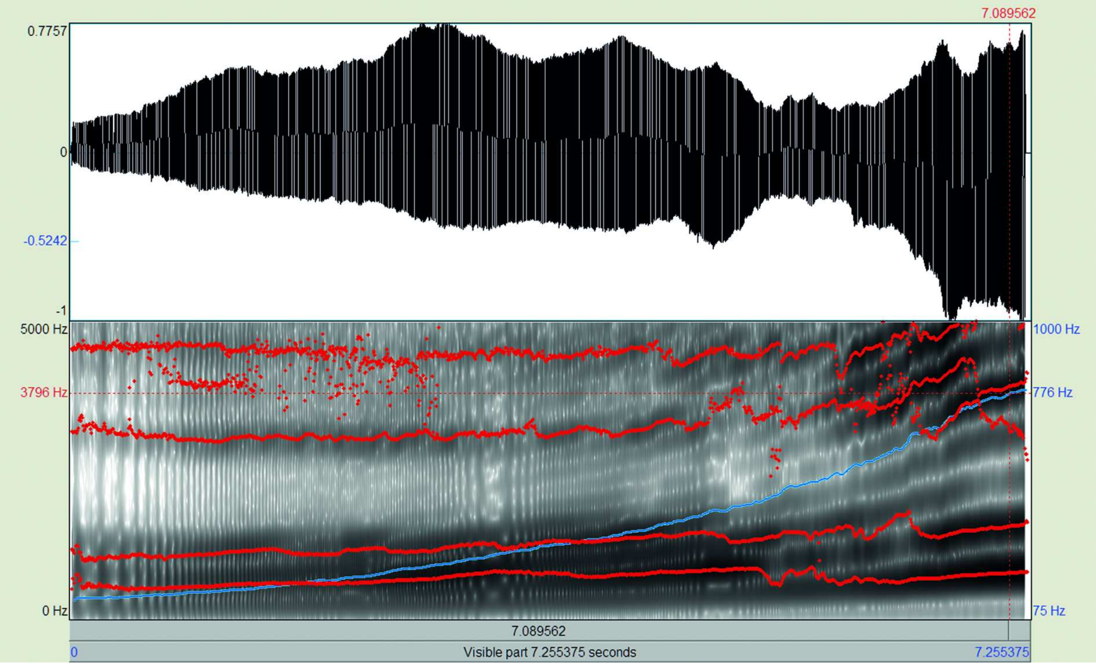

Fig. 1. The waveform and the spectrogram of the vowel /a/ pronounced by a female voice. The fundamental frequency contour F0, indicated by the blue line, changes in the 140–776 Hz range, the frequency of the first formant (F1, red line) increases from the 550 Hz mark, whereas the frequency of the second formant (F2, red line) increases from 1050 Hz to an indefinite value.

What also needs to be mentioned here is the perceptual normalization of the sender's vocal tract and elimination of individual variability, i.e. the ability to properly classify a phoneme produced by voices that exhibit significant acoustic differences. On the one hand, when we hear sounds with fairly similar F1 and F2, we identify them as different phonemes. On the other hand, there are male, female and child voices with very different F1 and F2 values of a phoneme, which is, however, identified as the same sound in a phonemic sense. The F0 frequency is a normalizing factor here: it plays a key role in the process of the perceptual normalization of the sender's vocal tract (Imiołczyk, 1991). Maurer *et al.* (2015) while investigating real vowel vocalization even argue that the differences in the position of the lowest formants ($<$ 1.5 kHz) disappear when the F0 value is identical for different groups of speakers: women, men and children. He puts forward the conclusion that these findings confirm the huge impact of F0 on the position of the lowest formants.

In the aforementioned work (Owsianny, 2001), the software formant synthesizer, modelled on Klatt's system (Klatt, 1980; Klatt, Klatt, 1990), was used to resynthesize the Polish vowels /i/, /y/, /e/, /a/, /o/, and /u/ produced by female and male speakers. These prototypical phonemes were modified by varying the F0 fundamental frequency in the range of ±1 octave with a 1/4 octave step in relation to the F0 course obtained for individual voices. The frequencies of the two lowest formants were also varied. 5,292 test signals were used. The listeners were tasked to identify vowels played in random sequence. The results show that any mismatch between the formant frequencies and the voice pitch results in a change in phonetic category. Female voices are more susceptible to F0 changes. Considerably decreased F0 values for female voices cause the vowel /y/ to be identified as /e/, /o/ is identified as /a/ while /u/ is identified as /o/. By contrast, increased F0 values for female voices result in the perceptual replacement of /e/ by /y/. The perceptual identification of the male /a/ as /o/ due to an increased F0 value or the replacement of the female /o/ by /a/ with a decreased F0 can easily be explained by the similarity of the formant frequencies of the male /a/ and female /o/. However, to explain the identification of the female /e/ as /y/ by more than a half of the listeners, it is necessary to invoke the effect of the centre of gravity (COG), as described by Chistovich *et al.* (1979), Traunmüller (1981), and Syrdal (1985), which provides a satisfactory account of this phenomenon. The effect consists in the frequency selectivity of the auditory system, i.e. the ability to hear particular components in a complex sound, which is linked with masking of the acoustic signal and the critical band theory (Jorasz, 1999). A different approach to this issue is taken by Johnson (1988a; 1988b), who focuses on the process of adjustment to the talker, in which F0 serves as an indicator of talker identification and vowel normalization is a result of the internal vowel space adjustment. Di Benedeto (1987) also notes that the perceptual boundaries between /i/ and /e/ in American English change due to F0 while Sundberg (1977), who studied singing voices, shows that vowel identification and their pitch level during singing can be maintained by lowering the mandible and raising mouth corners, which results in higher F1 values and the adjustment of the fundamental frequency to F1.

The present paper includes a series of studies on the perceptual identification of Polish natural oral vowels produced by various speakers and modified by a change in the frequency of the glottal tone F0. The results of the experiments are then compared with earlier results, which pertain, by contrast, to fully synthetic voices. In previous studies, many researchers restricted themselves only to determining the accuracy of vowel recognition, analysing a whole set of vowels or individual vowels in a specific language and their relation to F0. Moreover, they did not concentrate on erroneous identifications. Drawing on author's experiments and the ensuing conclusions, they were focused on misidentifications, i.e. perceptual replacement of an intentionally produced vowel by another one, which is misidentified by listeners due to the F0 change. Besides purely scientific values, such an approach can prove particularly useful in aiding the recognition of spontaneous and emotionally loaded speech in the automatic speech recognition systems (ASR) in the cases of big changes in the F0 parameter that could lead to the misidentification of vowels, which are the nucleus of the syllable and transfer suprasegmental information. Suggestions on the principles of the perception of Polish vowels can be used in the construction of acoustic models that are better adjusted to the real world of speech sounds and that will ensure a higher ratio of accurate recognition.

2. Description of experiments

Isolated Polish vowels were recorded in the following sequence: /i/, /y/, /e/, /a/, /o/, and /u/ with the sampling frequency of 44.1 kHz and a 16-bit depth, using the Shure Beta 58A dynamic supercardioid vocal microphone connected to the Zoom H4n recorder. This mode of recording ensured independence from the acoustic conditions of the surroundings and isolation from other sources of sound thanks to the microphone's directional characteristic. The vast majority of the recordings were made in a recording studio with a reverberation time of 0.5 seconds, the others in silent and strongly damped rooms. From among 57 subjects recordings, 31 voices were finally selected. The subjects were native speakers of Polish, aged 24 to 50 (the vast majority of them were young people). 21 subjects were female while 10 were male. The selection criteria were: vowel production quality and naturalness, recording quality (appropriate level and signal-to-noise ratio of more than 35 dB), vowel duration (from 200 to 300 ms), the interval between vowels much longer than the vowel duration itself (it totalled from 1.5 × to 4 ×, which ensured a small coarticulation effect), constant frequency of the glottal tone F0. All of these criteria were treated equally. The greater susceptibility of female voices to the perceptual shift was taken into account; thus, in the listening tests, the female voices outnumbered the male ones by a proportion of 2:1.

The recordings were edited, i.e. the intervals between the vowels were silenced and brought down to 0 dB. Then the recordings were modified by varying their fundamental frequency F0 in the range of ±1 octave with a 1/4 octave step. To this aim the PSOLA (Pitch Synchronous Over Lap-Add) algorithm was used (Mousa, 2010). PSOLA is generally regarded as a highly effective tool in F0 transformation as it mostly does not interfere with the spectral structure and provides a high quality of the modified signal (Kortekaas, Kohlrausch, 1997). Signal duration, formant frequencies and the general sound of each person's voice remained unchanged despite F0 modifications. In the experiment the "pitch fixer.praat" script was used. Written by Mark Antoniou of the MARCS Auditory Laboratories, and working under the PRAAT software (Boersma, Weenink, 2013), it automatically sets F0 to a predetermined value. The resynthesized vowels were checked for their acoustic properties; the analysis included F0 values, formant properties and the course of intensity. The F0 mean value for the female voices was 223.2 Hz (SD = 33 Hz), while for the male voices 123.4 Hz (SD = 18.6 Hz). Figure 2 shows sample waveforms and spectra for the vowel /a/ in the range of ±1 octave.

By means of the Sound Forge Pro 11.0 software, all recorded and transformed vowels were normalized in terms of sound intensity level. The effective levels of all vowels were brought to the same value of −18 dB RMS, which ensured acoustic and dynamic balance in terms of perceived stimulus pitch. Then the normalized vowel sequences were split into single phonemes and recorded. This yielded a total of 1,674 test signals saved in the .wav format. They came from 31 speakers (31 voices × 6 vowels × 9 F0 values). These signals were used in the listening experiments.

Assuming that a vowel is determined by the frequencies of the lowest two formants, each of them can be represented as a point in the F1 × F2 two-dimensional graph. Figures 3 and 4 present points that show the location of natural vowels pronounced by men and women, whose voices were used in the current experiment and the places of the occurrence (in the F1 × F2 coordinates) of synthetic vowels from the 2001 experiment (they are indicated by unfilled markers of the corresponding colour and asterisks). In the previous experiment, the first step was to select the natural vowel prototypes, and every prototypical vowel was

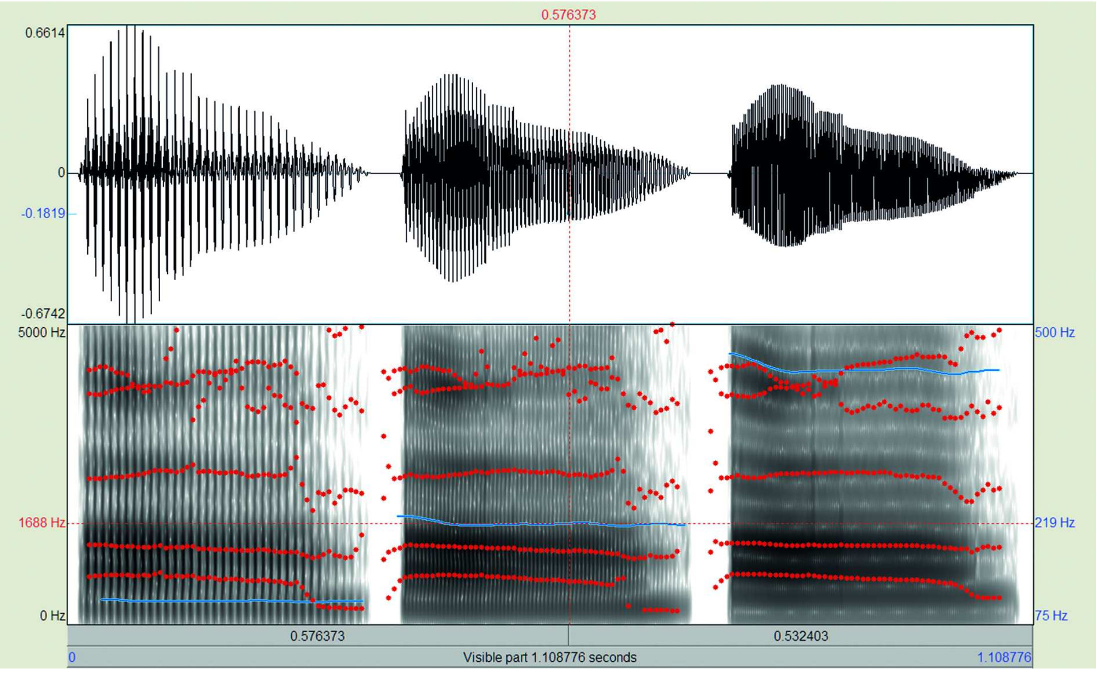

Fig. 2. The waveform and the spectrogram of the vowel /a/ pronounced by a female voice with the ΔF0 value of −1, 0, +1 octave. F0 transformation did not change the frequency of the formants. The fundamental frequency contour F0 was indicated by the blue line, formants contours by the red lines.

modified by changing the formant parameters, i.e. deviating the frequencies of the first and second formant by permanent and specified values. Specifically, the first formant (F1) was deviated by 50 Hz, 100 Hz and 150 Hz up and down from the real values whereas the second formant (F2) was modified by 150 Hz, 300 Hz and 450 Hz, respectively. The set of points belonging to one vowel formed a quantized area whose centre included the values of the two lowest formants of the prototype. Figures 3 and 4 show the areas of identification of the tested vowels produced by male and female voices from both experiments. It is worth noting that while in the current experiment the points represent different but natural values for real voices, the points from the previous experiment refer to the

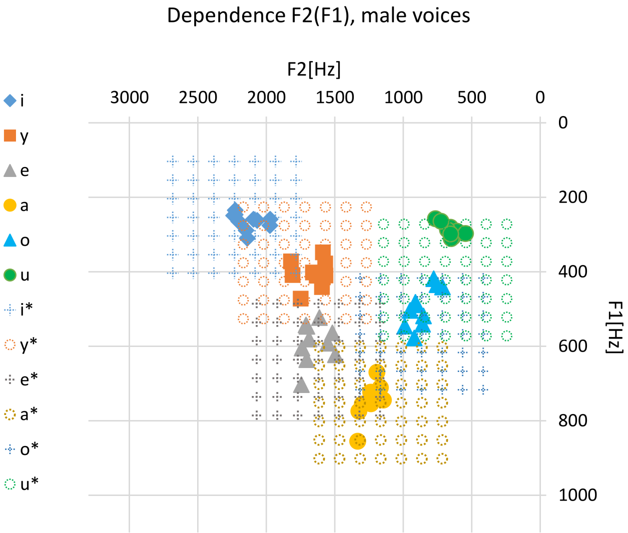

Fig. 3. The values of the two lowest formants of natural vowels pronounced by 10 males in the current experiment and the places of the occurrence (in the F1 × F2 coordinates) of synthetic vowels from the 2001 experiment (they are indicated by unfilled markers of the corresponding colour and asterisks).

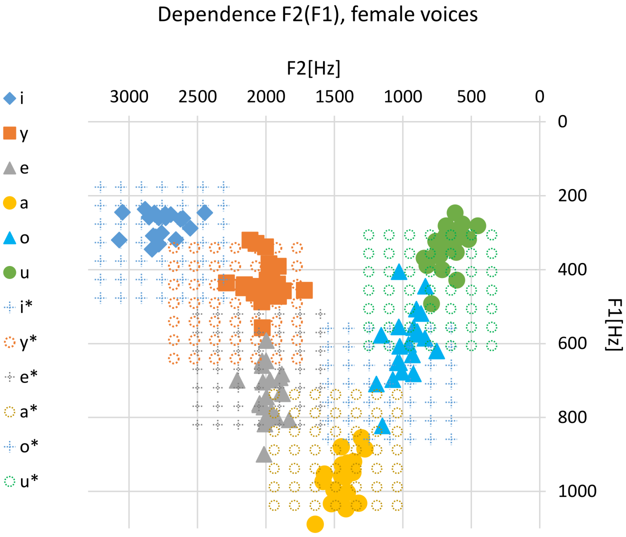

Fig. 4. The values of the two lowest formants of natural vowels pronounced by 21 females in the current experiment and the places of the occurrence (in the F1 × F2 coordinates) of synthetic vowels from the 2001 experiment (they are indicated by unfilled markers of the corresponding colour and asterisks).

hypothetical voices. It can also be said that, besides /o/ and /u/, the voices selected for the current experiments match the central positions of the earlier areas of vowel identification. The aforementioned phonemes, for both male and female, represented by the unfilled markers had too high F1 values.

2.1. Listening experiments

The experiments involved 20 otologically normal listeners aged 25–40 (11 females, 9 males). They were tasked to listen to vowels played in random sequence and assign them to one of the six phonetic categories. The listening session was composed of two 40-minute blocks. The stimuli were presented in random order separately for each listening session and for each of the subjects. For the purposes of the experiment, the PsychoPy open-source application was used. Written in the Python language (Peirce, 2007), this psychological software package enables users to monitor presented stimuli and save results to a .txt file. In the experiments, the listeners used the C/A USB DAC converter, the FiiO E10K amplifier, and the Sennheiser HD215 headphones. The experiments were conducted in the Poznań Supercomputing and Networking Center's recording studio. The six vowels were displayed on the monitor screen as potential answers. The listeners were asked to select one perceptually correct answer (with no repetitions) using a computer mouse or keyboard. The stimuli were presented at intervals of at least three seconds. The listeners could not use an extra category "unnatural sound" when answering the question about the kind of perceived vowel. If they could, they would most likely select this answer after hearing a modified vowel when exposed to quality changes. The main object of interest was potentially 'incorrect' answers indicating a perceptual shift.

The results of the experiment, i.e. the parameters of sent and received stimuli and reaction times, were recorded in a table, which enabled their multivariate analysis. The mean number of misidentified vowels (i.e. answers in which the listeners wrongly identified the stimulus by assigning it to a phonetic category other than the suggested prototype) was 117 out of 1,674 stimuli, which is 7% of the total. The incorrect answers were given far more frequently in the case of the female voices (6%) than the male ones (1%). Even if we consider the fact that the number of the prototypical female voices was twice that of the male ones, then the number of the misidentified female voices is still much higher. The proportions between the ratios of answers concerning the female and male voices are similar to the findings of the 2001 experiments. However, in the experiments conducted 17 years ago, probably due to the overlap of the areas of identification of individual vowels (Figs 3 and 4), the percentage of misidentified vowels was much higher and reached an average of 37%. A possible reason for this high rate of incorrect answers was worse quality of the vowel stimuli in the 2001 experiments: the sampling frequency was only 10 kHz while the depth was 8 bits and the vowel prototypes were differently defined (the vowels /o/ and /u/ had too high F1 values compared to the vowels identified in the current experiment even though those values were certainly within the range of the parameters of the Polish vowels). Last but not least, the vowels in the 2001 experiment were synthetic as they were produced in a formant synthesizer. Their quality, albeit worse than in the current realizations, did not raise any concerns then.

The total number of misidentified vowels results from both random and systematic errors, that is errors that are caused by the mode of presentation, unnaturalness of sound, contrast and coarticulation as well as the listeners' reactions to F0 induced vowel modifications, which were the aim of the perceptual experiment. The last kind of error which describes dependence on F0 is interesting from the perspective of the experiment's objective, that is determining the kinds of perceptual identification shifts from one vowel to another due to F0 as well as describing and accounting for this phenomenon.

With a view to testing this assumption, additional experiments were planned to be conducted. In these tests, the F0 range was to be increased up to ±2 octaves. In another experiment, CVC (consonant, vowel, consonant) pseudowords were to be used to test the identification of vowels in syllables. The new experiments were to be conducted solely on female voices. Table 1 provides an overview of all three experiments.

Table 1. Overview of the experiments.

	Experiment 1	Experiment 2	Experiment 3
Objective	Identification of 6 vowels with modified F0 values in the range of ±1 octave; comparison with the results of the 2001 experiment.	F0 range extensions: F0 in the range of ±2 octaves	Vowels in CVC structures; F0 in the range of 1.5/+2 octaves

No. of stimuli	21 female voices + 10 male voices;
	31 voices × 6 vowels × 9
	F0 = 1,674
		21 female voices;
		21 voices × 6 vowels × 7
		F0 = 882
			10 female voices;
			10 voices × 14 CVC pseudowords × 6
			F0 = 840

No. of listeners	20 (11 women + 9 men)	15 (7 women + 8 men)	12 (6 women + 6 men)

2.2. Results. Comparison with the results of the previous experiment

The most interesting finding relates to the results displayed in the graphs showing the identification of vowel pairs that come separately from the male and female voices between which changes in phonetic categories were observed in 2001. Phonetic change category was then defined as occurring when more than 50% of listeners have assigned a given vowel to another specific phonetic category. Phonetic categories perceptual shifts due to F0 changes, were observed then for male speakers (to a small degree) and female speakers. The charts in Fig. 5 show dependence of the vowel pairs identification on the F0 frequency for female speakers. The shape and direction of F0 changes observed for the 2001 experiment responses for this category of voices are consistent with the current experiment response characteristics. The 2001 experiment results, represented by dotted and dashed lines in the figure, are located close to the centre of the identification range, indicating that both vowels from a given pair were equally likely to be selected by the listeners. However, what needs to be emphasized is that the two experiments were based on different ways of vowel modification. In the earlier experiment,

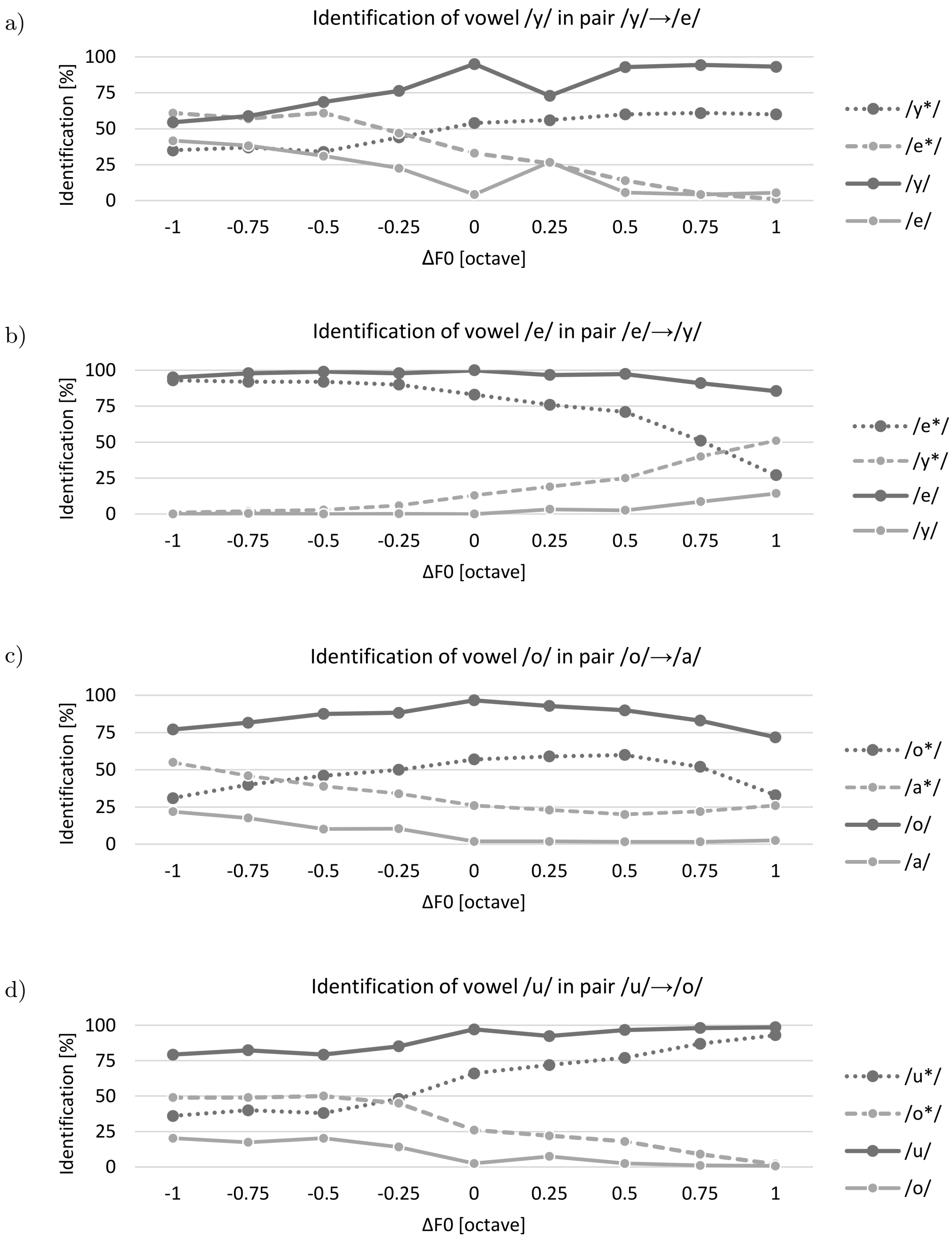

Fig. 5. Charts showing the rate of identification (in percentages) of the vowel pairs /**y**/→/e/, /**e**/→/y/, /**o**/→/a/, /**u**/→/o/ (produced by female speakers), within which shifts have been observed due to F0 changes. The dotted and dashed lines indicate the results of the 2001 experiment.

this was achieved by using one of the parameters of formant, spectral-parametric synthesis, which ensured the production of recurrent and unambiguous speech elements. In the current experiment, the PSOLA algorithm was used. The stimuli used in both listening experiments also differed in quality: in 2001 the sampling frequency was 10 kHz and the depth was 8 bits while in the current test the sampling frequency was 44.1 kHz and the depth was 16 bits, which is CD quality. The results of both experiments are consistent only with respect to the female vowels shown in Fig. 5. The courses of vowels produced by the male voices show some similarities to the data obtained earlier. However, the shifts towards another vowel are hardly noticeable (perceptual shift /**a**/→/o/). The charts showing the percentage rates for the identification of vowels pronounced by female speakers, that is /y/, /e/, /o/ and /u/ in the pairs /**y**/→/e/, /**e**/→/y/, /**o**/→/a/, /**u**/→/o/, between which shifts due to F0 changes have been observed, are indicated by the continuous line. The data for the 2001 experiments are indicated by the dotted and dashed line. The right arrow between the analysed vowel shifts shows a change in phonetic category due to increased or decreased F0 values. While the results have consistent shapes and direction of F0 changes, the curves obtained in the current experiment do not cross the 50% identification rate in the charts. Therefore, there is no indication of a change in phonetic category, as defined above. Only the pair /**y**/→/e/ (that is a change in the identification of the female vowel /y/ into the male /e/ due to decreasing F0 by −1 octave) is close to meeting this requirement. The rate of identification of /y/ as /e/ with F0 decreased by −1 octave is only 41.7%, failing thus to meet the 50% requirement. For the same F0 value, the identification rate for /y/ as /y/ is slightly lower, at 54.5%. In the remaining cases, the listeners identified /y/ as /u/ (2.4%) and /o/ (1.4%). The sum of all percentages is 100%, which is equivalent to 315 listener responses. As for a specific pair, in this case /**y**/→/e/, the first chart (Fig. 6a) shows the identification of the vowel /y/ (the upper part of the chart) and the identification of the vowel /y/ recognized as /e/ (the lower part of the chart) and does not include the other phonemes indicated by the listeners (i.e. /u/ and /o/), the upper and lower lines for a given pair are not symmetrical.

Even though the courses in the charts of Fig. 5 that show the pairs /**e**/→/y/, /**o**/→/a/, /**u**/→/o/ are similar in shape, it cannot be argued that due to F0 modifications performed by PSOLA, natural vowels undergo a change in phonetic category. What actually happens is a perceptual shift in vowel identification as opposed to the vowel shift on the F1 × F2 plane mentioned in the introductory section of this paper. If the F0 range was continually increased, it would likely result in a change in phonetic category.

It is assumed that the closeness of the areas of identification has an impact on the incorrect assignment of vowels to a specific phonetic category. Even though in the 2001 experiment, there was a greater overlap in the areas of the male vowels than in the case of the female voices, the shift effect occurred more frequently in the female voices. The vowel misidentification rate for the female voices is around seven times higher compared to the male ones in both the 2001 experiment and in Experiment 1. What is the reason for this? Real and natural voices are almost entirely separated from each other on the F1 × F2 plane, as is shown in Figs 3 and 4; yet there are identification shifts, especially in the female voices. This special case of female voices could be explained by the fact that by increasing the frequency of the glottal tone we transfer them in a natural way to the area of child voices, and by decreasing the frequency we bring them to the area of male voices. Obviously, the sheer multiplying/dividing F0 values is not sufficient for creating natural sounding child or male voices (Owsianny, 1994). "High fundamental frequency causes the harmonics of female voices to appear further apart from each other. This, in turn, facilitates the processes of auditory integration of harmonics in the vicinity of F1 into one 'centre of gravity' (Chistovich *et al.*, 1979; Syrdal, 1985), which can lead to the auditory shift of the first formant and, in result, to phonetic category shift." (Owsianny, 2001, p. 203). This conclusion can easily be proven considering the identification of the vowel /e/ as /y/ due to increased F0 values in the 2001 experiment (indicated by the dotted and dashed lines in Fig. 5b). By doubling the frequency of the glottal tone F0, an initially high F1–F0 value for /e/ (about 4.5 barks) decreases to the value far lower than 3 barks (units of a subjective pitch sound scale), which enables a perceptual shift F1 towards lower frequency values, and makes the new pattern of formants similar to the vowel /y/. Doubling the F0 value for the current experiment proved insufficient: there was no change in phonetic category; yet a certain perceptual shift has been observed. However, by applying the same theory, it is difficult to describe the perceptual shift of the vowel /y/ to /e/ with a decreased F0 value by 1 octave (Fig. 5a), primarily due to the high density of the harmonics. After all, the curves for the 2001 case and the current one are surprisingly consistent.

Worth noting is the lack of symmetry in the identification of the vowel /o/ (Fig. 5c) as well as in the case of other pairs, which results from the fact that the listeners identified a specific vowel as different from the other vowel in a given pair. Here are the data for the shift of /o/ to /a/ due to the modified fundamental frequency ΔF = +1 octave: in 71.9% of the responses the vowel /o/ was identified as /o/, in 24.3% as /u/, in 2.6% as /a/, in 0.7% as /i/, and in 0.5% as /e/. Hence, there is no symmetry between the identification

Table 2. Mean opinion score and standard deviation as a function of frequency change F0 obtained as a result of testing the quality/naturalness of vowels presented to listeners in experiment 1. The study was conducted using the subjective MOS test method. A five-point scale was used: from 1 (poor quality) to 5 (perfect quality).

ΔF0 [octave]	−1.00	−0.75	−0.50	−0.25	0.00	0.25	0.50	0.75	+1.00
MOS score	2.92	3.44	3.86	4.03	4.39	4.12	3.80	3.57	3.37
SD	0.70	0.50	0.40	0.33	0.35	0.39	0.47	0.54	0.60

of /o/ and /a/. This case is also interesting due to double, two-sided changes in identification: with ΔF = −1 octave /o/ is identified as /a/ in 21.9% of cases, with ΔF = +1 octave /o/ is identified as /a/ only in 2.6% of cases, but in 24.3% of cases as /u/.

An interesting dependence (which is intuitively completely obvious) has been observed and confirmed. For all dependences of identification on F0 (Fig. 5) displayed in the charts, the curves of identification of vowel pairs for natural vowels (continuous lines) for the unchanged fundamental frequency ΔF0 = 0 are always very high and equal to 100% or a little less for an investigated vowel, and close to 0 for a vowel towards which identification is shifted due to F0. Interestingly, such a phenomenon is not observed in the case of synthetic voices in the 2001 experiment (dotted lines). It turns out then that the fundamental frequency F0, which is natural for a particular voice, receives special and exceptional treatment as it is always correctly recognized by the listeners. It is proven by the fact that a specific vowel represented by the two lowest formants needs to be linked with adequate F0 (Imiołczyk, 1991), which listeners can correctly recognize. A natural pitch of a vowel pronounced by the speaker is privileged by listeners.

Using the analysis of variance ANOVA, the impact of F0 on listener reaction times (the time lapse between stimulus generation and indication of the correct answer) has been investigated. At the significance level $p = 0.05$, there are no grounds for rejecting the zero hypothesis that listener reaction times are independent of different values of F0, which is an independent variable. F0 change has no impact on listener reaction times during the identification of specific vowels.

Tests of quality, naturalness were performed using MOS (mean opinion score) for signals from experiment 1. A five-point scale was used: from 1 (poor quality) to 5 (perfect quality). Prior to the experiments, the listeners were adequately instructed and prepared (natural vowels with different pitches produced by female speakers were presented). However, out of the 7 listeners taking part in the experiment, the results of only five were further analysed. The results of the listeners who misidentified the real vocalizations for the un-deviated F0 were omitted.

A rather uncommon vowel "roughness" was observed with the fundamental frequency F0 decreased by −1 octave. This feature was confirmed by the results of a test of stimulus quality (below). The phonemes with F0 decreased by −1 octave received by far the most negative ratings (over 58% of all answers in this category, which is only 0.8% of total answers). For F0 = −0.75 octaves, the 1 point rating was given in 22% of answers while in the case of the high F0 = +1 octave, in 3.7% of responses. It should be added that, despite some unnaturalness, the phonemes with extreme F0 deviations were still easily identifiable in terms of phonetic category evaluation.

2.3. F0 range extensions. Experiment 2

The data for the identification of vowel pairs produced by the female speakers in experiment 1 (they confirm a perceptual vowel shift due to F0 changes in the ±1 octave range but do not indicate a change in phonetic category) have been subjected to a twice-larger F0 change to test the claim about a change in phonetic category in female voices due to F0 modifications. Such an F0 change is common in spontaneous speech, especially in the production of highly emotionally-loaded speech. The glottal tone frequency was changed in the ±2 octave range with a $1/2$ octave step. Models of six Polish vowels produced by 21 female speakers from the previous experiment were used. The stimuli were obtained by changing vowel sequences using the PSOLA algorithm. The vowels sequences were then normalized in terms of sound intensity level to the value of −18 dB RMS and split into single phonemes. The entire set of the stimuli totalled 882 items (21 voices × 6 vowels × 7 F0 values) was evaluated by 15 listeners (7 females and 8 males). They, like in the previous experiment, identified individual vowels presented in random order. The mean number of incorrect answers for each listener was 162, which accounted for 18% of all responses. The incorrect answers for specific vowels were as follows: /i/ (3%), /a/ (6.4%), /u/ (8.9%), /e/ (22%), /y/ (26.8%), /o/ (33%). The charts in Fig. 6 presenting the rate of identification of the vowels /y/, /e/, /o/, and /u/ as a function of F0 show the distribution of incorrect answers. Even though the phonetic category did not change in every vowel pairs under investigation, only in the case of the vowel /e/ for ΔF0 = +2 octaves, the change in phonetic category, /e/ into /y/, is prominent (Fig. 6b). For the same F0 value change, the rate of identification for /y/ markedly decreases to 54% (Fig. 6a) while for the vowel /o/ even to 17.5% (Fig. 6c). In this last case,

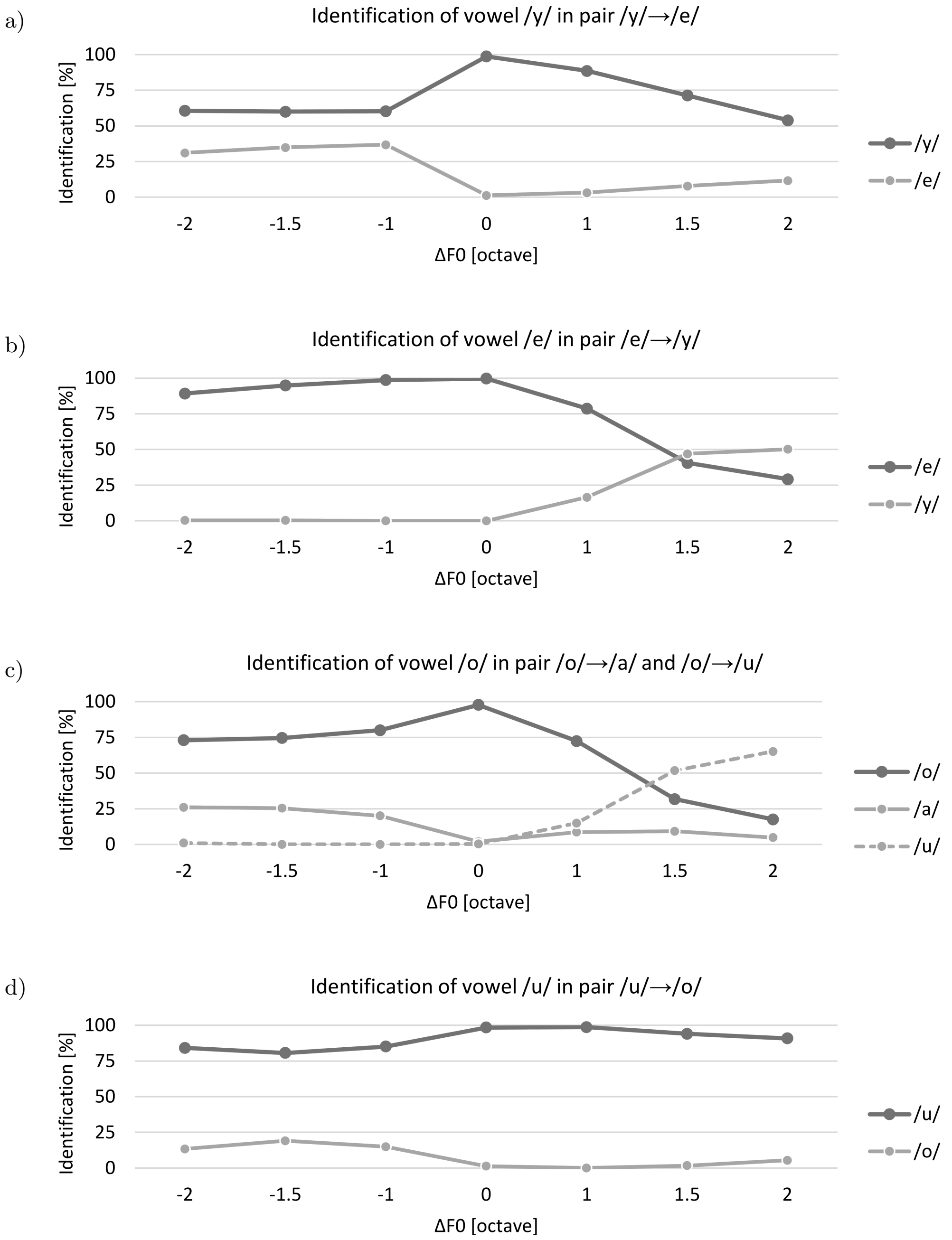

Fig. 6. Charts showing the rate of identification (in percentages) of the vowel pairs /**y**/→/e/, /**e**/→/y/, /**o**/→/a/, and /**u**/→/o/ (produced by female speakers), within which shifts have been observed due to F0 changes in the range of ±2 octaves.

the identification of /o/ spreads over other phonemes and is identified by the listeners as /i/: (1.6%), /y/ (2.2%), /e/ (8.9%), /a/ (4.8%), /o/ (17.5%), and /u/ (65.1%). These results, especially the last one, explain the low identification rate for the vowel /o/ with F0 twice increased compared to the natural value. The vowel /o/ produced by 21 female voices was in such circumstances identified as /u/ by 65.1% of the listeners. Accordingly, the phonetic category has changed, from /o/ to /u/, as is shown in Figure 6c. Hence, an additional curve has been added to show the identification rate for the vowel /u/. The notation /o/→/u/ indicates that with F0 values equal to 1.5 and 2 octaves, in more than 50% of cases, the listeners identified the vowel /o/ with high F0 as /u/. The shape of the curves from Fig. 6 is in fact consistent with the dependencies of identification as a function of F0 observed with smaller F0 value changes from the first experiment. Increasing the range of F0 changes did not, however, cause expected changes in phonetic category of /y/ to /e/ with decreased F0 values (Fig. 6a), of /o/ to /a/ (Fig. 6c) or of /u/ to /o/ (Fig. 6d) with decreased F0. In all likelihood, with such large F0 changes, the quality of transformations of vowels generated by the PSOLA algorithm proved to be insufficient. In view of the result of the 2001 experiment, an expected change in phonetic category due to increased F0 values occurred in the perceptual shift of /e/ to /y/ (Fig. 6b). An unexpected result was a significant decrease in the identification rate for the vowel /o/ in favour of /u/ with F0 values increased up to 1.5 and 2 octaves (Fig. 6c).

2.4. Investigating CVC structures. Experiment 3

The empirical material in this experiment consisted of 840 CVC structures (produced by female speakers) with a fricative consonant in the onset and the nasal consonant /m/ in the rime with different F0 values. The vowel phonemes were represented by /e/ and /y/, between which shifts due to F0 changes were observed. All of the utterances were pseudowords composed in accordance with Polish phonotactics and did not constitute lexical units. Fricative consonants were thus selected to follow the above rules. For this reason, /s'/ and /z'/ were omitted as these two fricatives cannot be followed by the VC combination that includes the nasal /m/ in the rime. As a result, a total of the following 14 utterances were created: /hem/, /hym/, /sem/, /sym/, /fem/, /fym/, /Sem/, /Sym/, /zem/, /zym/, /vem/, /vym/, /Zem/, /Zym/. The list of utterances, written in standard orthography, was read aloud by 10 female speakers and then recorded and saved in the .wav format in CD quality. Prior to the experiment, it was assumed that F0 values would be changed by ±2 octaves with a 1/2 octave step. In actuality, however, it proved impossible to create a set of stimuli with ΔF0 = −2 octaves out of the available inventory of patterns (using the PSOLA algorithm). As was the case with the previous experiments, the pseudowords were normalized in terms of sound intensity level to the value of −18 dB RMS and split into single utterances. The complete set of stimuli totalled 840 test signals (10 female voices × 14 CVC pseudowords × 6 F0 values) which were to be identified in the listening experiment.

Twelve people took part in the experiment: 6 females and 6 males. The number of incorrect answers, i.e. responses with the misidentified middle vowel, ranged across the participants from 87 to 155, giving an average of 116, which was 14% of all responses. The standard deviation was 23. The analysis of the results excluded errors caused by the misidentification of the fricative in the onset. The syllable-final /m/ remained unchanged in all of the utterances under investigation. If the listener misidentified a consonant but correctly identified the vowel, such cases were not regarded as identification errors.

The results are displayed in Fig. 7, in the charts that show the identification rates for the pairs of CVC pseudowords (including /y/ and /e/, with varying F0 values) produced by female speakers. The identification of /y/ does not show strong dependence on the F0 frequency (Fig. 7a) or similarity to course of the identification of an isolated vowel from Fig. 6a. What we do observe, however, is that the course of the identification of the vowel /e/ in a CVC structure as a function of F0 (Fig. 7b) is similar to the course of the identification of an isolated vowel from (Fig. 6b). Moreover,

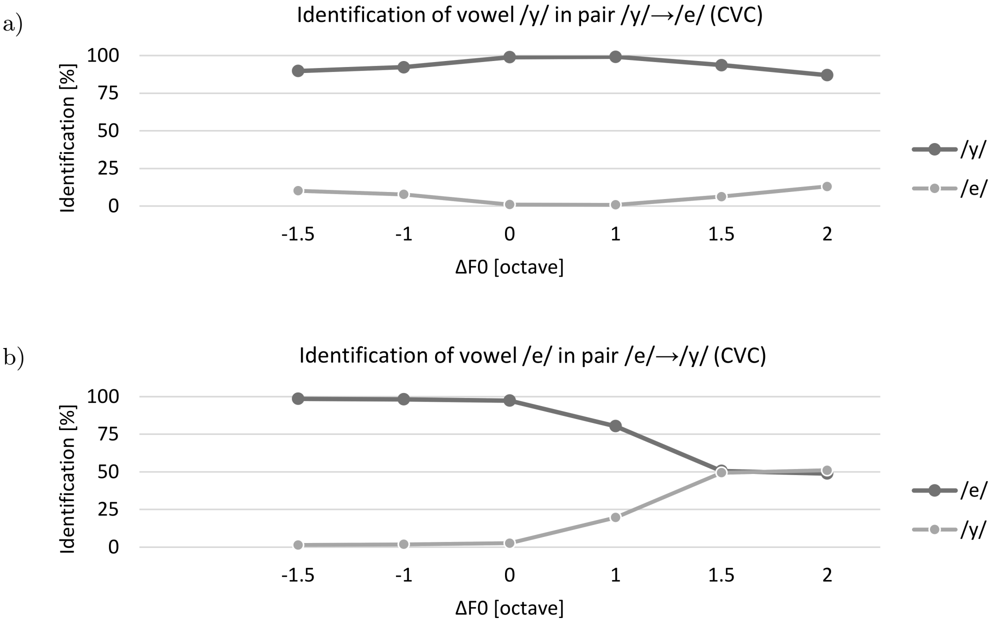

Fig. 7. Charts showing the rate of identification (in percentages) of the pairs of CVC pseudowords including the vowels /**y**/→/e/and /**e**/→/y/ (produced by female speakers), within which shifts have been observed due to F0 changes in the range of +2/ – 1.5 octaves.

there is a change in phonetic category: /e/ changes into /y/ with ΔF0 equal to ca. +1.5 octaves. A very similar threshold limit has been recorded for the isolated vowel /e/. Deviation from the glottal tone frequency by a small positive value in relation to the investigated models results in a perceptual shift towards the vowel /y/ whereas as a result of the deviation by +1.5 octaves, over 50% of the listeners misidentify the vowel /e/ in a CVC structure.

The impact of the fricatives used in the experiment on the identification of the vowels /y/ and /e/ in the CVC structure was examined. A statistical evaluation of the results has been conducted, using the one-way ANOVA. It was found that the influence of individual fricatives (independent variable) was statistically insignificant in relation to the mean identification (dependent variable) at the significance level $p = 0.05$ for both vowels /y/ and /e/.

3. Discussion

The experiments on the identification of Polish natural oral vowels (occurring in isolation and in CVC structures) produced by various speakers and modified by F0 frequency changes show the significance of this parameter in the case of female voices. These voices were thus used in the majority of the experiments. The obtained results were then compared with the results of the experiments on synthetic voices. The courses of the curves for the identification of natural and synthetic vowels as a function of the fundamental frequency F0 for female voices show consistency in shape and direction of F0 changes. However, not always did this result in a change in phonetic category, i.e. the case when more than 50% of the listeners identified a vowel as a different phoneme (Fig. 5). Such a change occurred only when F0 values were more deviated from the +1 octave range (Figs 6b, 6c, 7b). Sometimes, as was the case with the change of the female vowel /y/ to /e/, greater deviation of the F0 parameter towards lower frequencies from experiment 2 (Fig. 6a) although they confirm the results of experiment 1 (Fig. 5a) (considering the lack of common data for ΔF = −2, −1.5, +0.25, +0.5, +0.75 octave), it did not lead to a change in the phonetic category, which may indicate insufficient quality of the assessed stimuli or different and listener-specific reactions to the same set of stimuli. Another factor at play may be the impact of the experience from the previous listening experiments. However, it needs to be emphasized that the direction of the perceptual shift has remained unchanged (for ΔF0 = −1 from Fig. 5a a change in phonetic category almost occurred whereas for the same value of deviated F0 shown in Fig. 6a, it did not occur). Sometimes, it was enough to increase the glottal tone frequency by 0.5 octaves to cause an expected perceptual change of /e/ to /y/ (based on the course of the identification curves for synthetic voices) for natural voices at ΔF0 = +1.5 octaves (Fig. 6b). And it even happened that a similar increase in F0 values in the case of the perceptual identification of the vowel /o/ caused an unexpected but very spectacular change in phonetic category of /o/ to /u/ (Fig. 6c). As regards the vowels /e/ and /y/ in CVC structures, the results did not confirm a change of /y/ to /e/ due to decreased F0 (Fig. 7a), which did occur in experiment 2 in the case of isolated vowels (Fig. 6a). However, the reverse change of /e/ to /y/ with increased F0 values to +1.5 octaves is undeniable (Fig. 7b) and consistent with the analogous change for isolated vowels (Fig. 6b).

The courses of the identification curves as a function of F0 for synthetic and natural voices were similar even though the percentage of misidentified vowels was markedly different, i.e. four times higher than in the case of synthetic voices. However, it should be noted that the arbitrarily selected area of occurrence of a synthetic vowel in the F1 × F2 plane was significantly larger; moreover, there was a partial overlap between these areas.

In both current experiments a relatively huge discrepancy was reported between the number of misidentified vowels perceived by the experiment participants and the high standard deviation (SD) values. In the first experiment, the SD was 49 while in the second one – 44 (these figures were independent of the listeners' gender and age). The key factor was experience in listening to and identifying sounds. The best results, in terms of the smallest number of errors, were found among the young listeners, and the participants who had some professional background in the assessment and annotation of recordings.

The quality of the vowels transformed by the PSOLA algorithm, especially for low and very high F0 values, showed many imperfections and serious errors. Some of them could not be removed, which is why it was decided not to transform CVC pseudowords with ΔF0 = −2 octaves.

Changes in phonetic category were also studied by Hirahara and Kato (1992) in their paper titled *The effect of F0 on vowel identification.* In their research, they analysed male voices, which were evaluated by separate groups of male and female listeners. In the experiments, F0 was not decreased below its base value for vowels. What is important, however, is that the researchers increased the values of four formants along with the fundamental frequency. They very consistently applied the subjective scale of sound pitch (measured in barks) to describe frequencies. They presented a schematic diagram of the course of vowel identification that included theories of hearing and the tonotopic organization of the cochlea. The proved the existence of the following changes in phonetic category for male voices due to increased F0: /e/ to /i/ or /u/, /a/ to /o/, and /o/ to /u/. As the authors argue, "the F0 affects the vowel quality in some stimuli even when the formant set remains unchanged" (Hirahara, Kato, 1992, p. 108). But, interestingly, simultaneous changes in F0 values and specific formants along the bark scale did not result in vowel quality changes.

It can thus be concluded that the F0 frequency serves an important normalizing function in the perception of speech sounds. Whether we hear a male, female or child voice, we select an appropriate scale and find a suitable pattern of reinforcements of specific formants frequencies, which identify a particular vowel (Johnson, 1988a; 1988b). Of special significance is also the centre of gravity (COG) hypothesis put forward by Chistovich *et al.* (1979), Traunmüller (1981) and Syrdal (1985), which binds all formants with a 'comb' of harmonics that extend due to increased F0. This effect of the auditory integration of F0 and harmonics is largely responsible for the perceptual change in phonetic category for mid high F0 values and occurs across a sequence of neighbouring sounds, from close to open. A suitable F0 frequency for a particular voice defines it and imposes limitations on the range of possible changes for each individual speaker. An F0 value which is unadjusted to a particular voice has a bad impact on vowel identification as it can change the perceptual phonetic category of particular phonemes, against the speaker's intention. As already mentioned, the obtained results show that for the unchanged F0 values (ΔF0 = 0), the identification rate for all of the vowels used in the listening experiments was close to 100%. Deviation from the F0 value resulted in decreased identification rates (sometimes by over 50%), and led to a perceptual shift or a change in phonetic category. It seems impossible that there is no defensive mechanism, based on auditory feedback, that would prevent the misidentification process. Perhaps increased formant values along with increased F0 for a specific voice (Hirahara, Kato, 1992; Chladkova *et al.*, 2009; Maurer *et al.*, 2015) play such a role and eliminate at least some errors caused by changes in the F0 frequency. Expanding the F1 × F2 vowel space in the case of female voices, observed by Diehl *et al.* (1996) and raised by the spectral undersampling hypothesis, provides additional evidence for this finding. Similar proposals are put forward by Assamann and Nearey (2008), who argue that vowel identification is more precise when F0 and the spectral envelope scale factor increase or decrease together, compared to the situation when only one of these factors is subject to change. However, there is an important limitation: the physical dimensions of the vocal tract and the range of articulator movements (especially the tongue movements). Hence, for very large F0 changes during singing, for example, vocalists need to search for another solution and resort to various tricks to attain the intended effect of maintaining vowel identification and their volume level (Sundberg, 1977). Such an approach explains why female voices are more easily affected by F0 changes and why the perceptual vowel shifts and changes in phonetic category are more frequent in female voices. Their vowel space is larger by nature and it cannot grow any further as F0 values increase, which leads to distorted identification. An experiment on the identification of child voices as a function of the F0 fundamental frequency should provide insights into the dependence between the F0 frequency and the spectral envelope (formant frequencies). Of special interest are limitations of vowel space in child voices. Thus, there are plans to conduct further perceptual studies of natural child voices with F0 values changed by the digital signal processing technique. Another idea is to use the results of the 2001 experiment in an attempt (through statistical methods) to investigate the under-researched problem of a temporary vowel shift, i.e. naturally increased F1 and F2 formant frequency values due to an increased voice pitch for a speaker's gender (Fig. 1). As a result, this research can contribute to providing additional arguments for the compensatory impact of formant frequencies on the improvement of the perceptual identification of vowels changed by the glottal tone frequency.

4. Conclusion

The paper showed the impact of voice pitch on the perceptual identification of Polish oral vowels, including isolated synthetic phonemes and natural phonemes which have been modified by F0 changes based on the PSOLA algorithm. The study also investigated vowel identification in CVC structures. All of these measurements were made with unchanged frequencies of formants in relation to F0 changes. The paper presented quantitative findings for the Polish language regarding all possible kinds of perceptual shifts from one phonetic category to another as a function of voice pitch. It is female voices that are most susceptible to such changes. The most likely shifts for female voices due to increased F0 values are /e/ to /y/ and /o/ to /u/ or /o/ to /a/.

Acknowledgment

Part of the present study was carried out within the framework of the project "Automatic analysis of phonetic convergence in speech technology systems" supported by the Polish National Science Centre (no.: 2014/14/M/HS2/00631).

References

1. Assamann P.F., Nearey T.M. (2008), *Identyfication of frequency schifted vowels*, The Journal of the Acoustical Society of America, **124**, 5, 3203–3212.

2. Boersma P., Weenink D. (2013), *PRAAT: doing phonetics by computer* [Computer program]. Version 5.3.59, retrieved December 11, 2014 from http://www.praat.org.

3. Carlson R., Fant G., Granström B. (1975), *Two-formant models, pitch, and vowel perception*, [in:] G. Fant, M.A.A. Tatham [Eds.], *Auditory analysis and perception of speech*, Academic Press, London, pp. 55–82.

4. Chistovich L.A., Sheikin R.L., Lublinskaya V.V. (1979), *"Centers of gravity" and spectral peaks as the determinants of vowel quality*, [in:] B. Lindblom, S. Öhman [Eds.], *Frontiers of Speech Communication Research*, Academic Press, London, pp. 143–157.

5. Chladkova K., Boersma P., Podlipsky V.J. (2009), *On-line formant shifting as a function of F0*, Proceedings of the INTERSPEECH 2009 Conference, pp. 464–467, Brighton, UK.

6. Di Benedetto M.G. (1994), *Acoustic and perceptual evidence of a complex relation between F1 and F0 in determining vowel height*, Journal of Phonetics, **22**, 205–224.

7. Diehl R.L., Lindblom B., Hoemeke K.A., Fahey R.P. (1996), *On explaining certain male-female differences in the phonetic realization of vowel categories*, Journal of Phonetics, **24**, 187–208.

8. Długosz-Kurczabowa K., Dubisz S. (2006), *Historical grammar of Polish language* [in Polish: *Gramatyka historyczna języka polskiego*], Wydawnictwo Uniwersytetu Warszawskiego, Warszawa, pp. 96, 129.

9. Fant G. (1960), *Acoustic theory of speech production*, Mouton, Hague.

10. Hirahara T., Kato H. (1992), *The effect of F0 on vowel identification*, [in:] *Speech perception, production and linguistic structure*, Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka [Eds.], Ohmsha, Tokyo, pp. 89–112.

11. Imiołczyk J. (1991), *Determination of perceptual boundaries between the male female and child's voices in isolated synthetic polish vowels*, Archives of Acoustics, **16**, 2, 305–323.

12. Jassem W. (1992), *Acoustic-phonetic variability of polish vowels*, Archives of Acoustics, **17**, 2, 217–233.

13. Johnson K. (1988a), *F0 normalization and adjusting to talker*, Research on Speech Perception, Progress Report 14, pp. 237–258.

14. Johnson K. (1988b), *Intonational context and F0 normalization*, Research on Speech Perception, Progress Report 14, pp. 81–108.

15. Jorasz U. (1999), *Selectivity of the auditory system*, Adam Mickiewicz University Press, Poznań, pp. 38–51.

16. Klatt D.H., Klatt L.C. (1990), *Analysis, synthesis, and perception of voice quality variations among female and male talkers*, Journal of the Acoustical Society of America, **87**, 2, 820–857.

17. Klatt D.H. (1980), *Software for a cascade/parallel formant synthesizer*, Journal of the Acoustical Society of America, **67**, 971–995.

18. Kortekaas R.W.L., Kohlrausch A. (1997), *Psychoacoustical evaluation of the pitch-synchronous overlapand-add speech-waveform manipulation technique using single-formant stimuli*, Journal of the Acoustical Society of America, **101**, 4, 2202–2213.

19. Maurer D., Suter H., Friedrichs D., Dellwo V. (2015), *Gender and age differences in vowel-related formant patterns: What happens if men, women, and children produce vowels on different and on similar F0?*, Journal of the Acoustical Society of America, **137**, 4, 2416–2416.

20. Meister E., Werner S. (2009), *Vowel category perception affected by microdurational variations*, Proceedings of the INTERSPEECH 2009 Conference, pp. 388–391, Brighton, UK.

21. Mousa A. (2010), *Voice conversion using pitch shifting algorithm by time stretching with PSOLA and resampling*, Journal of Electrical Engineering, **61**, 1, 57–61.

22. Obrębowski A. (2008), *Vocal organ and its importance in social communication* [in Polish: *Narząd głosu i jego znaczenie w komunikacji społecznej*], Wydawnictwo Naukowe Uniwersytetu Medycznego w Poznaniu.

23. Owsianny M. (1994), *The synthesis of female voices using a software synthesizer*, Archives of Acoustics, **19**, 2, 185–199.

24. Owsianny M. (1995), *The effect of voice pitch on the perception of synthetic Polish vowels*, Proceedings of the 4th European Conference on Speech Communication and Technology – EUROSPEECH'95, pp. 945–948, Madrid, Spain.

25. Owsianny M. (2001), *Interaction between vocalic quality and fundamental frequency in the perception of Polish vowels*, Proceedings of the PROSODY 2000 Conference, Speech Recognition and Synthesis, pp. 197–204, Kraków, Poland.

26. Peirce J.W. (2007), *PsychoPy – Psychophysics software in Python*, Journal of Neuroscience Methods, **162**, 1–2, 8–13.

27. Peterson G.E., Barney H.L. (1952), *Control methods used in a study of the vowels*, Journal of the Acoustical Society of America, **24**, 175–184.

28. Sundberg J. (1977), *The Acoustics of the singing voice*, Scientific American, **236**, 3, 82–4, 86, 88–91.

29. Syrdal A.K. (1985), *Aspects of a model of the auditory representation of American English vowels*, Speech Communication, **4**, 121–135.

30. Traunmüller H. (1981), *Perceptual dimension of openness in vowels*, Journal of the Acoustical Society of America, **69**, 5, 1465–1475.