Part III Experiences and Observations

The third part of the main text formulates several hypotheses about the actual relationship between vowel sounds, sound spectra and formant patterns. These hypotheses refer to the recordings mentioned in the first part of the introduction and to related analyses and observations.

7 Unsystematic Correspondence between Vowels, Patterns of Relative Spectral Energy Maxima and Formant Patterns

7.1 Inconstant Number of Vowel-Specific Relative Spectral Energy Maxima and Incongruence of Vowel-Specific Formant Patterns

As discussed in Section 3.1, sounds of back vowels and of /a–ɑ/ can exhibit only one relative spectral energy maximum within their vowel-specific frequency range ≤ 1.5 kHz (≤ 2 kHz for some sounds of /a/), in contrast to other sounds of the same vowels, which have two such maxima. Consequently, the number of vowel-specific energy maxima is inconstant.

The spectral envelopes and formant patterns of such vowel sounds cannot in all cases be interpreted as "formant merging": examples of sound pairs of back vowels can be observed for which both sounds exhibit the lowest spectral envelope peak at a similar frequency level, but only one of them has a pronounced second envelope peak within the frequency range mentioned. Then, the first spectral envelope peak of both sounds corresponds to the vowel quality in question, whereas the second spectral envelope peak may be linked to an additional "colouring" of that sound. However, it plays a marginal role in vowel perception and, in such a case, does not posses vowel-differentiating value.

For both sounds of such sound pairs, formant analyses using current methods may reveal two lower formants. However, calculating F2 for the first sound of the respective sound pair mentioned, exhibiting only one lower spectral envelope peak, may prove highly contingent on the number of filters chosen, above all for sounds of children. In addition, its amplitude can be very low and its bandwidth can be very large, that is, far beyond reference values as given in the literature.

With regard to front vowels, the frequency of observable second envelope peaks, and with them also calculated F2, can vary strongly. Because of this, there are examples of sound pairs of front vowels for which the second envelope peak and calculated F2 of one sound approaches or even exceeds the third envelope peak and calculated F3 of the other sound. (Such observations in general relate to sounds of speakers of different speaker groups, which are produced at similar fundamental frequencies. However, this can also be observed for the sounds of speakers of the same speaker group.)

Thus, it is not possible to designate a standard number of consecutive relative spectral energy maxima related to delimited frequency ranges that represent any given vowel. The same holds true for formants, although it is less obvious. There are also formant patterns of sounds of single vowels whose reciprocal correspondence of single formants is open to discussion.

The number of vowel-specific relative spectral energy maxima is inconstant, and formant patterns are incongruent in some cases.

7.2 Partial Lack of Manifestation of Vowel-Specific Relative Spectral Energy Maxima

In their vowel-specific range of the spectrum ≤ 1.5 kHz, sounds of back vowels and of /a–ɑ/ produced at fundamental frequencies ≤ 350 Hz can exhibit series of harmonics with consistent, quasi-identical amplitudes. These vowel-specific parts of harmonic spectra seem to be “flat”, lacking any clearly distinctive relative energy maxima. Of special interest in this respect are the sounds of /a, ɑ, ɔ, o/ in cases where the amplitudes of the first three to five harmonics are not markedly different.

In their vowel-specific range of the spectrum ≥ 1.5 kHz, sounds of front vowels produced at fundamental frequencies ≤ 350 Hz can also exhibit series of harmonics with consistent, quasi-identical amplitudes. Thus, what applies to back vowels and to /a–ɑ/ for their entire vowel-specific frequency range also applies to front vowels for the higher part of their vowel-specific frequency range.

In addition, cases of such vowel-specific, “flat” spectral portions also exist for sounds produced at fundamental frequencies > 350 kHz, even if, in relation to the large frequency spacing of the harmonics, this generally remains limited to the sounds of the vowels /i, e, ɛ, a, ɑ/. For certain fundamental frequencies of the sounds of /ɔ, o/, the first two harmonics can exhibit equal amplitudes.

Also worth mentioning in this context are the sounds of back vowels and of /a–ɑ/, which exhibit continuously decreasing amplitudes in the vowel-specific lower frequency range. In the spectra of these sounds, the first harmonic generally forms the actual spectral maximum.

Thus, the set of problems concerning a formulation of a general relationship between the perceived vowel quality and its physical representation based on a certain number of relative spectral energy maxima is again extended.

Spectral envelope maxima, as described in the literature, are not a precondition for the physical representation of vowels.

The relationship between “flat”, vowel-specific parts of sound spectra and calculated formant frequencies using current methods of analysis cannot be described in simple and general terms. The same holds true for the relationship between continuously decreasing amplitudes of the harmonics in the vowel-specific lower frequency range and calculated formant patterns. Therefore, the issue is left open to discussion here. However, it has to be considered as an additional methodological problem of formant analysis.

7.3 Addition: Resynthesis and Synthesis

Inconstancy in the number of vowel-specific relative spectral energy maxima, possible incongruence of formant patterns and vowel sounds with “flat” or decreasing vowel-specific spectrum portions can be replicated using resynthesis.

The same also applies to formant patterns or harmonic spectra not derived directly from natural vowel sounds.

8 Lack of Correspondence between Vowels and Patterns of Relative Spectral Energy Maxima or Formant Patterns

8.1 Dependence of Vowel-Specific, Relative Spectral Energy Maxima and Lower Formants ≤ 1.5 kHz on Fundamental Frequency

If investigated empirically and systematically, it becomes evident that the first spectral envelope peak—if it exists—and the first calculated formant of vowel sounds often depend on fundamental frequency.

For a range of fundamental frequencies ≤ 350 Hz for which formant analysis is not critical in principle, this dependence is particularly evident in the sounds of the vowels /e, ø, o/ at fundamental frequencies in the range of 200 Hz to 350 Hz.

For a range of fundamental frequencies > 350 Hz, this dependence is, above all, indicated in sounds of the vowels /i, y, u/, because the first harmonic generally exhibits the highest amplitude; thus, the lowest spectral peak rises with increasing fundamental frequency.

In addition, such a dependence can also be observed for the second formant for cases of sounds of back vowels.

For sounds of /ɛ/ and of /a–ɑ/, however, indications of a dependence of F1 on fundamental frequency may prove to be weak and corresponding observations may require a comparison of sounds with a very extended vocal range.

Moreover, the observation of a dependence of F1 on fundamental frequency is not only related to frequency ranges of the latter and vowel qualities but also to single speakers and their phonation characteristics, including vocal effort. (Note that marked differences in the vocal effort of vowel production have a substantial effect on spectral peaks and calculated formant frequencies, and this effect has to be taken into account when investigating the relationship between F0, spectral peaks and formants.) But although the indications for the dependence discussed here prove to be unsystematic, the findings of intelligible vowel sounds at fundamental frequencies > 500 Hz (see next chapter) and of formant pattern ambiguity (see Chapter 9) force us to relate the lower spectral peaks and the lower formants to fundamental frequency.

The possible relationship between fundamental frequency and higher vowel-specific spectral envelope peaks or formants > 1.5 kHz for sounds of front vowels is left open here for discussion.

These assertions hold true for vowel sounds produced by one and the same speaker. Thus, they apply to vowels and their physical representation.

In this respect, what is of particular importance is the observation that the dependence of lower spectral envelope peaks and lower formants ≤ 1.5 kHz does not represent a phenomenon generally related to "oversinging" the first formant of a vowel: most importantly, the shifts of F1 in the sounds of the vowels /e, ø, o/ can already be observed at fundamental frequencies substantially below the corresponding statistical values for F1 as given in the literature for sounds produced in citation-form words. Moreover, given a range of fundamental frequencies of c. 200–350 Hz, the shifts of F1 for the sounds of the vowels /e, ø, o/ are in many cases much more pronounced than for the sounds of the vowels /i, y, u/, although, for the former, the literature gives significantly higher statistical values for F1 than for the latter.

Also of particular importance—and foreshadowing formant pattern ambiguity of vowel sounds (see Chapter 9)—is the observation that, in many cases of sounds of a vowel produced by a single speaker, the shifts of F1 in relation to fundamental frequency exceed the F1 differences of two neighbouring vowels as given in formant statistics for a corresponding speaker group (for speakers with corresponding vocal-tract size). In line with this, the shifts mentioned also exceed speaker-group differences in F1 for that same vowel as given in the format statistics mentioned.

Vowel-specific relative spectral energy maxima ≤ 1.5 kHz (if determinable) and calculated vowel-specific formant patterns (if methodologically substantiated) are dependent on fundamental frequency.

8.2 Vowel Perception at Fundamental Frequencies above Statistical Values of the First-Formant Frequency

Speakers possessing a large vocal range and good phonation and articulation are able to form the sounds of the vowels /i, y, e, ø, ɛ, a, o, u/ in a recognisable and distinguishable way up to a fundamental frequency of c. 700–800 Hz. Such sounds can be readily experienced up to a fundamental frequency of c. 600 Hz because they occur frequently in everyday speech, in particular among children and women. However, these sounds can also be evidenced for men using "falsetto".

Speakers possessing excellent vocal abilities are even able to form the sounds of the corner vowels /i, a, u/ in a clearly recognisable and distinguishable way up to a fundamental frequency of c. 800–1000 Hz. (Ongoing research also indicates that other vowels, too, are intelligible in this vocal range.)

Correspondingly, the respective sound spectra exhibit vowel-specific differences, even if these have to be described other than in terms of spectral envelopes and formant patterns, for example in terms of vowel-specific configurations in the levels of the harmonics (see below, Sections 13.2 and 13.3).

Note that a fundamental frequency of 700 Hz lies above the statistical F1 values given for sounds of all long German vowels produced by women or men, except for /a/ of women. A fundamental frequency of 800–1000 Hz even lies above the statistical F1 values for all long German vowels, for both women and men (see Section 2.2).

The vowel quality of sounds produced at fundamental frequencies above statistical values of the vowel-related first-formant frequency is intelligible in principle.

The possibility of such vowel production and perception contradicts the designation of established, statistically determined formant patterns as "vowel-specific" patterns, irrespective of the methodological problems of determining envelope peaks and formant frequencies. At the same time, vowel perception and discrimination at such high fundamental frequencies confirms that lower spectral energy maxima (if determinable) and lower formants (if methodically substantiated) depend on fundamental frequency.

The vowel quality of sounds of back vowels and of /a–ɑ/ produced at fundamental frequencies > 500 Hz can be physically represented solely in terms of the first two or three harmonics and their amplitudes. This accentuates the basic problem of assuming that relative spectral energy maxima, that is, envelope peaks in closely delimited frequency ranges, are a pervasive physical characteristic of the sound of a vowel.

Here, the question of the maximal fundamental frequency up to which all vowels of any given language can in principle be produced in a recognisable way is left open for discussion.

8.3 "Inversions" of Relative Spectral Energy Maxima and Minima and "Inverse" Formant Patterns in Sounds of Individual Vowels

Given that spectral envelope peaks ≤ 1.5 kHz (if determinable) depend on fundamental frequency, pairs of sounds of a back vowel produced at different fundamental frequencies can exhibit "inverse" relative spectral maxima and minima in the form of "inverse" spectral envelope curves ≤ 1.5 kHz without any change in vowel perception: whereas we see a relative minimum in the spectrum for one sound, we may observe a spectral maximum for the other, and vice versa. The same holds true for comparisons between the respective calculated filter curves and formant patterns (if methodologically substantiated): where for one sound, the filter curve exhibits a relative minimum, for another sound, the curve may exhibit a maximum, and vice versa.

In the case of some front vowels, such "inversions" can also be observed for the higher vowel-specific frequency range, even if the question of the relationship between such "inversions" and fundamental frequency variation is left open here.

This observation reaffirms the lack of a general correspondence between vowels, vowel-specific spectral envelope curves and corresponding formant patterns.

> With regard to vowel-specific frequency ranges, the spectral envelope curves of two sounds of the same vowel produced at two different fundamental frequencies can exhibit "inverse" behaviour. The same holds true for formant patterns.

8.4 Addition: Whispered Vowel Sounds, Fundamental-Frequency Dependence of Vowel-Specific Spectral Characteristics and "Inversions"

As discussed in Section 5.5, formant statistics indicate increased vowel-specific formant frequencies F1 and F2 for whispered sounds when compared to voiced sounds. However, according to the corresponding recording procedures of the comparative investigations, this only applies to the lower range of fundamental frequency of the voiced sounds produced in citation-form words, comparable to relaxed speech in an enclosed space.

Given that a whispered sound exhibits higher first and second formants than a voiced sound of the same vowel and given that the latter's fundamental frequency is gradually increased during its production, then in many cases it is possible to determine a certain fundamental frequency for which F1 and F2 of the whispered and voiced sound correspond with each other.

Whether this represents an actual rule is left open here.

If the fundamental frequency of a voiced sound is increased further, then there will be cases in which F1 or F1–F2 of the whispered sound are lower than F1 or F1–F2 of the voiced sound.

In any event, the general statement that whispered sounds exhibit fundamentally higher vowel-specific formant patterns than voiced sounds does not apply.

Over the course of such experimentation, cases involving comparisons between whispered and voiced sounds exhibiting the described "inversions" may also be found.

8.5 Addition: Resynthesis and Synthesis

All the above aspects of the lack of correspondence between vowels and patterns of relative spectral energy maxima or formant patterns, discussed in relation to natural vowel sounds, can be evaluated and replicated using resynthesis.

The same holds true for resynthesis at fundamental frequencies > 350 Hz related directly to the harmonic spectra of natural vowel sounds.

The same also applies to synthesis involving formant patterns or harmonic spectra not derived directly from natural vowel sounds.

9 Ambiguous Correspondence between Vowels and Patterns of Relative Spectral Energy Maxima or Formant Patterns or Complete Spectral Envelopes

9.1 Ambiguous Patterns of Relative Spectral Energy Maxima and Ambiguous Formant Patterns

All these reflections and observations come down to the conjecture that two sounds of two different vowels, produced at two different fundamental frequencies, can exhibit quasi-identical relative spectral energy maxima and quasi-identical formant patterns within their vowel-specific frequency range. Indeed, many patterns of spectral envelope peaks and formants prove to be ambiguous empirically. As such, they often physically represent two (or even several) different vowels.

> In many cases the patterns of relative spectral energy maxima do not prove to be vowel specific, but ambiguous. The same holds true for formant patterns.

This observation becomes particularly evident when comparing vowel sounds for their entire range of fundamental frequencies for which vowels are recognisable and distinguishable and when evaluating the correspondences between relative spectral energy maxima and minima also in a direct comparison of harmonic spectra, aside from determining spectral envelopes and formant frequencies.

9.2 Ambiguous Spectral Envelopes

In certain cases, this ambiguity also concerns the entire course of the spectral envelope.

> Spectral envelopes can be equally ambiguous.

9.3 Ambiguity and Individual Vowels

For all German vowels discussed here, there are cases of sounds with ambiguous patterns of relative spectral energy maxima or with ambiguous formant patterns within the respective vowel-specific frequency ranges.

To what extent this is also true for complete spectral envelopes is left open for discussion.

If vowel sounds are compared for their entire range of fundamental frequencies for which vowels are recognisable and distinguishable and if a possible correspondence of relative spectral energy maxima and minima is evaluated in a direct comparison of harmonic spectra, then, the above ambiguity can be observed not only for sounds of neighbouring vowel pairs but also for other sound pairs and sometimes for sounds of more than two different vowels. This holds particularly true when comparing sounds produced by all of the three age- and gender-related speaker groups.

The ambiguity described is not limited to only a part of the vowels or to neighbouring vowel pairs, and it can affect more than two vowels simultaneously.

The question of whether there are sounds of certain vowels that exhibit strict vowel-specific patterns of relative spectral energy maxima and strict vowel-specific formant patterns, which cannot be found in sounds of any other vowel—for example for sounds of /a/—is left open for further discussion.

9.4 Addition: Resynthesis and Synthesis

The ambiguity discussed above in relation to natural vowel sounds can be evaluated and replicated using resynthesis.

The same also applies to synthesis involving formant patterns or harmonic spectra not derived directly from natural vowel sounds.

10 Lack of Correspondence between Patterns of Relative Spectral Energy Maxima or Formant Patterns and Speaker Groups or Vocal-Tract Sizes

10.1 Similar Patterns of Relative Spectral Maxima and Similar Formant Patterns ≤ 1.5 kHz for Different Speaker Groups or Different Vocal-Tract Sizes

If sounds of a vowel are produced at equal fundamental frequencies by children, women and men, and if these sounds perceptually correspond with each other not only in terms of their general attribution to a vowel quality but also in terms of the respective "vowel-colour" variant—which makes for the greatest possible correspondence as regards perception—then, empirically, both the relative spectral energy maxima (if determinable) and the formant patterns (if methodically substantiated) often prove to be similar in the lower frequency range ≤ 1.5 kHz, apart from possible differences due to the different parameter settings involved in formant analysis. Expected age- and gender-related spectral differences decrease or disappear if the fundamental frequency of the utterances correspond for children, women and men.

Further, for sounds of back vowels and sounds produced by men at higher fundamental frequencies than women, it follows that the sounds of men (at higher F0) may exhibit higher relative spectral energy maxima (if determinable) and higher F1 or even F1–F2 patterns (if methodically substantiated) than the sounds of women (on lower F0), as holds true for F1 of front vowels. The same may also occur in a corresponding comparison of sounds of adults and children.

No statements are made here on /a–ɑ/ since our observations do net yet allow for general formulations for all sounds of /a–ɑ/ (see Section 8.1).

Thus, the question arises whether the lower range of the vowel spectrum mentioned is indeed dependent on age- and gender-related speaker groups, that is, on vocal-tract size. In the literature, this lower frequency range is referred to as being entirely vowel specific for all back vowels and, concerning F1, vowel specific for all other vowels.

In any event, the general statement that the sounds produced by children exhibit the highest, the sounds of women intermediate and the sounds of men the lowest patterns of vowel-specific relative spectral energy maxima and formant frequencies does not apply.

Within the frequency range of ≤ 1.5 kHz, vowel-specific patterns of relative spectral energy maxima (if determinable) and formant patterns (if methodically substantiated) often prove to be empirically independent of the age- and gender-related speaker group, that is, the vocal-tract size. Given strict perceptual correspondences, then, differences refer directly to the differences in fundamental frequency.

As mentioned, the possible relationship between fundamental frequencies and higher vowel-specific spectral envelope peaks or formants for sounds of front vowels is left open for discussion. In the present context, this also concerns the question of whether or not higher frequency ranges are in principal specific to vocal-tract sizes.

10.2 The Dichotomy of the Vowel Spectrum

As mentioned repeatedly, while the dependence of vowel-specific spectral characteristics and formants on fundamental frequency for the lower frequency range ≤ 1.5 kHz is easily understandable and reproducible empirically, this is not the case for the higher frequency range. At the same time, lower spectral ranges and lower formant frequencies are not generally specific to speaker groups and vocal-tract sizes. Whether this is also the case for higher spectral ranges and formant frequencies is still in question. Thus, the spectrum of a vowel sound needs a twofold rather than a uniform consideration.

The spectrum of a vowel proves to be dichotomous.

In this context, with regard to the sounds of front vowels, it is particularly important to consider that, in certain cases, higher relative spectral energy maxima (if determinable) and higher formants (if methodically substantiated) > 2 kHz may be simultaneously related to vowel identity and perceived speaker group: differences in this higher frequency range can often be observed for sounds of a front vowel produced by children, women and men if the speakers form these sounds at similar fundamental frequencies, even if there is no such difference found in the lower frequency range.

However, it is left open for further investigation whether this is also the case if men imitate so-called “female voices” or if adults imitate “children’s voices”.

10.3 Addition: Whispered Vowel Sounds and Speaker Groups or Vocal-Tract Sizes

No results of comparative studies of formant patterns for whispered vowel sounds of children, women and men have been published to date that have obtained a reference status as is the case for reference statistics of voiced vowel sounds referred to in Part II. However, the studies that compare whispered sounds of different speaker groups (limited in number and generally not including all vowels of a language) refer to corresponding differences between formant patterns.

Notwithstanding the reflections and comments made so far, these differences can be understood as an indication of a general relationship between patterns of relative spectral energy maxima and formant patterns on the one hand, and speaker groups, that is, average vocal-tract sizes on the other, including the lower frequency ranges.

This aspect and its significance regarding the relationship between vowels and related spectral characteristics is left open to discussion here and needs to be clarified and discussed elsewhere.

10.4 Addition: Vowel Imitations by Birds

Sounds of animals imitating utterances of humans are also of primary importance in the discussion of vowel sounds, related spectral characteristics, formant patterns, perceived speaker groups and vocal-tract sizes.

Fundamental in this respect is the question of how birds are able to imitate human sounds despite lacking the means of phonation and articulation—in particular, a corresponding vocal-tract resonance.

According to our own preliminary examination of vowel imitation by common hill myna birds who excel at such mimicry (results unpublished, although some clear examples are given in the Materials section), we conclude the following: if these birds imitate words, and if individual imitated vowel sounds are isolated as sound fragments in a way that they possess a quasi-static character in terms of quasi-static spectral characteristics (above all, that transitions are excluded), then vowel perception and a distinction of such sounds by humans is possible. For part of these sound fragments, complete F1–F2–F3 formant patterns comparable to patterns given for human sounds can be interpreted. For the remaining fragments, only a partial correspondence in formant patterns can be observed. (However, this statement must be relativised: strictly speaking, any calculation of vowel-related formant patterns of bird sounds is methodically unsubstantiated; see below.)

The fact that birds are able to imitate human vowel sounds with vowel-specific spectral characteristics and formant patterns comparable to those of humans contradicts, in its turn, a strict correspondence between the spectral characteristics of the produced sound and vocal-tract resonance. The same holds true for a strict correspondence between spectral characteristics of the produced sound and vocal-tract size. Consequently, any critical investigation and discussion of vowels must focus on the possibility that the same sound characteristics can be produced under substantially different physical and physiological conditions.

Besides, if birds are able to mimic human utterances, they must be able to perceptually differentiate different vocal sounds. However, their perception cannot rely on any sensomotoric and conceptual experience of vowel production comparable to the experience of humans. Thus, it can be speculated that their perception relies on a more "abstract" acoustic "form" of the vowel sound. (Such speculation would meet the claim that a phenomenological approach to the physical representation of vowels is needed; see Part V.)

10.5 Addition: Resynthesis and Synthesis

Again, the lack of a general correspondence between patterns of relative spectral energy maxima or formant patterns and speaker groups or vocal-tract sizes can be evaluated and replicated using resynthesis and synthesis.

11 Lack of Correlation between Methodological Limitations of Formant Determination and Limitations of Vowel Perception

11.1 Vowel Perception at Fundamental Frequencies > 350 Hz

As discussed in Section 8.2, recognisable vowels can be produced at fundamental frequencies substantially exceeding the critical limit above which formants can no longer be reliably determined for methodological reasons.

> Vowel perception is maintained for sounds at fundamental frequencies > 350 Hz. Yet, for these middle and higher fundamental frequency ranges, formant pattern estimation is questionable for methodological reasons. Thus, the methodological limitation of determining formant patterns of vowel sounds at fundamental frequencies > 350 Hz does not coincide with impaired vowel intelligibility.

Consequently, formulating a general theory of the physical representation of vowels based on formant patterns proves to be critical due to the related methodological limitations.

11.2 Lack of Correspondence between Methodological Problems of Formant Pattern Estimation at Fundamental Frequencies ≤ 350 Hz and Impaired Vowel Perception

Vowel sounds produced at fundamental frequencies ≤ 350 Hz, for which the estimation of formant patterns proves questionable for reasons other than fundamental frequency—for instance, if expected relative spectral energy maxima are "missing" or if vowel-related parts of a spectrum are "flat"—are not less recognisable than vowel sounds for which formant pattern estimation may be said to be unproblematic.

> Methodological problems regarding the determination of formant patterns of vowel sounds at fundamental frequencies ≤ 350 Hz do not necessarily coincide with impaired vowel intelligibility.

11.3 Addition: Lack of Methodological Basis of Determining Formant Patterns for Vowel Mimicry by Birds

Given the prevailing methodological standards, strictly speaking, the imitation of human vowel sounds by birds cannot be studied in terms of formant patterns. As explained in Section 6.3, formant calculation requires parameter settings for the frequency range and the maximum number of filters used in the analysis in relation to a specific vocal-tract size. Birds, however, have no vocal tract comparable to that of humans. Hence, it is impossible to determine how many filters should be used in analysing a vowel-like sound produced by a bird to determine vowel-specific formants.

Thus, in a first step, comparisons between the utterances of humans and birds must be based on a direct comparison of the respective spectra and must relate to the interpretation of observable relative spectral energy maxima. However, in a subsequent step, formant analysis double-checked by resynthesis may be applied even if methodically unsubstantiated, in order to foster the discussion.

Again, this methodological limitation of mimicry analysis does not coincide with a principal difficulty to identify the imitated vowel sounds involved.