Timbre Solfege and Auditory Profile Analysis

Jan ŻERA
Warsaw Institute of Technology, Faculty of Electronics and Information Technology,
Institute of Radioelectronics and Multimedia Technology, Nowowiejska 15/19,
00-665 Warszawa, j.zera@ire.pw.edu.pl

Abstract

A course in auditory evaluation of sound, called *Timbre Solfege*, was developed by a team of researchers headed by Professor Andrzej Rakowski, at the Music Acoustics Laboratory, Chopin University of Music. A large part of the course, taught at the Department of Sound Engineering, has been focused on the detection and identification of timbre changes produced by formants and by other kind of sound spectrum modifications. Detecting formants in sound recordings is an auditory task that has much in common with auditory profile analysis, an area of research initiated and developed in psychoacoustics by Professor David M. Green, exploring the fundamentals of detection of changes in the sound spectrum envelope shape, independently of the differences in loudness between the sounds. The purpose of this study is an attempt to relate the results of the *Timbre Solfege* sound evaluation drills to the theory of the auditory profile analysis.

Keywords: sound spectrum, auditory profile analysis, formants, detection and discrimination

1. Introduction

The purpose of this paper is to compare detection thresholds of a formant created in the spectrum of noise (just noticeable formant amplitude) and discrimination thresholds of formant amplitude (just noticeable change in formant amplitude) with postulates of auditory profile analysis [1]. The data for such a comparison were taken from research studies oriented towards the development of auditory skills needed for professional practice in sound recording. The studies were part of a research program concerned with *Timbre Solfege*, a course aimed at perceptual and analytical evaluation of sound taught at the Chopin University of Music [2]. Auditory profile analysis is an area of research in psychoacoustics focused on the ability of discerning changes in the spectral envelope shape of complex acoustic signals. Thus, the measurements of formant detection and discrimination conducted within the *Timbre Solfege* research studies are both auditory tasks that can be examined with the use of the profile analysis methodology.

2. The *Timbre Solfege* course

The *Timbre Solfege* course (in Polish: *solfeż barwy*) was developed at the Music Acoustics Laboratory, Chopin University of Music, by a team of researchers led by late Professor Andrzej Rakowski. The course was introduced to the sound engineering curriculum in 1977 and since then its program and teaching methodology have been constantly expanded and modified. The course became a major four-semester practicum offered to students to develop various auditory skills, including detection and identification of formants, discrimination of pitch and loudness, recognition of various dimensions of timbre, and assessment of sound quality. The course program also includes exercises intended to develop the skill of analysing auditory spatial information and understanding the principles of auditory image formation. Detailed description of the course program and teaching methods can be found elsewhere [2-7]. It has been proved in progress tests conducted throughout the course that *Timbre Solfege* markedly improves the students' sensitivity to timbre changes and memory for timbre [7].

Along with teaching, the *Timbre Solfege* course creates a basis for research in timbre perception, such as investigations of the detection [8] and discrimination [9] of formant amplitude, discrimination of the formant frequency [10] and bandwidth [11]. Among these research areas, investigations of formant detection and discrimination reported by Letowski and Rogala [8], and Rogala and Śliwka [9] will be considered later in this paper as compatible with the scope and structure of auditory profile analysis [1].

Letowski and Rogala [8] measured detection thresholds for a single 1/3-octave formant created in white noise, and similarly modified the spectra of two natural sounds – a recording of tubular bells and a sample of pop music recording (Figure 1a). Results of this experiment, shown in Figure 1b, present the formant levels corresponding to 75% correct responses obtained with the use of paired comparison method [8]. The formant level needed for formant detection in white noise (solid line in Figure 1) was large at low frequencies, amounting to nearly 11 dB at a 63-Hz frequency of the formant band. In high frequency range, the level of the just detectable formant decreased to about 2 dB. The results obtained for natural sounds are non-monotonic, and their discussion will be postponed as irrelevant to the purpose of current study.

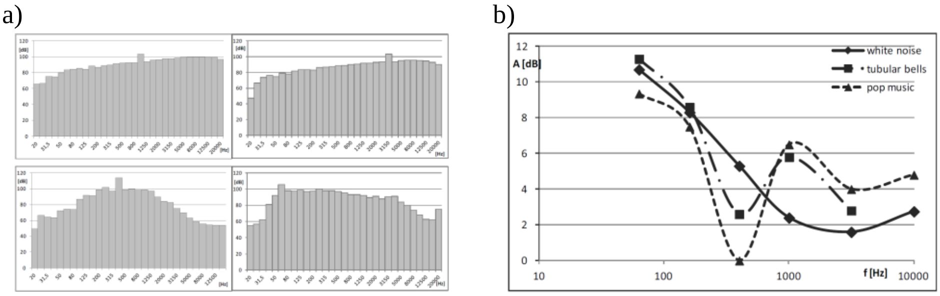

Figure 1. a) Upper panels: long-time average spectrum of white noise with added 1/3-octave formants at 1000 Hz (left) and 3150 Hz (right); lower panels: 400-Hz formant in tubular bells (left) and 63-Hz formant in pop music (right) spectra. b) Detection thresholds for formants created in spectra shown in a). Figures from ref. [8]

In the second experiment to be discussed here, reported by Rogala and Śliwka [9], formant discrimination thresholds were measured as a function of initial formant level and formant frequency. The initial level of a 1/3-octave formant imposed on wideband pink noise was either 3 or 12 dB. Using the *oddity* stimulus presentation method (modified *3-alternative forced choice* paradigm), a psychometric function for the detection of a change in the formant level was determined. The discrimination threshold was defined at 67% correct responses (middle point of the psychometric function). Average formant level discrimination thresholds, obtained for trained musicians and non-musicians, taken from reference [9], are presented in Figure 2.

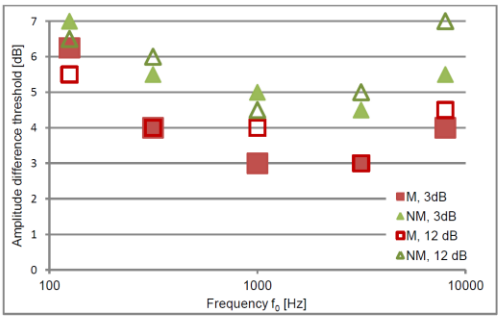

Figure 2. Discrimination thresholds of formant levels for musicians (M) and non-musicians (NM). Initial formant levels: 3 and 12 dB. Figure from reference [9]

The thresholds plotted in Figure 2, measured for centre noise band frequencies of 125, 400, 1000, 3150 and 8000 Hz, show a clear bowl shape, regardless of the formant initial level (3 or 12 dB) and the subjects' experience level. This bowl shape of threshold, favouring the 1-kHz frequency region, is characteristic of some results of auditory profile analysis and will be discussed in section 4. It is also apparent that discrimination thresholds for 3- and 12-dB formants are very similar, and differ by less than 1 dB. A constant discrimination threshold, in decibels, obtained for different formant levels indicates that Weber's Law holds for formant discrimination. This finding also suggests that the listeners may not perform spectral profile analysis, but relay on successive intensity comparisons in a frequency channel corresponding to the formant frequency. The thresholds measured for inexperienced subjects (NM group) are about 2–3 dB higher than those of experienced subjects (M group).

The results shown in Figures 1 and 2 will be compared in section 4 with the postulates of classical auditory profile analysis described by Green [1]. This will be done to examine, whether in the listening strategy in formant detection and discrimination tasks is based on successive signal comparisons in a single frequency channel or on simultaneous, across-channel comparisons.

3. Auditory profile analysis

Auditory profile analysis is a research direction introduced by David M. Green and his associates as a formalized approach to the auditory analysis of the spectral envelope profile of sound (for summary of early but substantial work see ref. [1]). The idea is to extend traditional intensity discrimination between successive sound samples by simultaneous comparison of intensity changes in different parts of the sound spectrum. The latter is simply discrimination of changes in the spectral envelope and thus the timbre of sound. The basic task consists in detection of a change in the intensity of a single component in a multicomponent complex. With this respect, profile analysis is similar to the detection of 1/3-octave formant in noise, earlier described in this paper. During a single presentation of a sound sample (*e.g.*, the signal) in a profile analysis task the subject simultaneously compares the energy level resulting from an amplitude change in a given frequency channel with energy levels in all the other channels, centred at different frequencies. Subsequently, he/she makes a similar simultaneous comparison during the presentation of a second sound sample (*e.g.,* the standard). Detection of an amplitude increment is done by a comparison of different frequency regions instead of comparison of an energy change in a single frequency channel, in successive presentations, as is the case during loudness discrimination.

Although in everyday listening to spectrally complex sounds both successive and simultaneous comparisons are performed at the same time, proper measurement in laboratory settings requires to separate these two factors one from another. Roving of the overall level is used in profile analysis experiments to force the simultaneous across frequency channels comparisons by making successive level comparisons ineffective. From trial to trial, the overall signal level is randomly set with equal probability within 10-, 20- (a typical value) or even 40-dB range, usually in 1-dB steps. Such a roving of level prevents the subject from performing successive intensity comparisons, and forces him/her to use only across-spectral comparisons as a cue for detection of a spectral change in a signal. The roving range is directly related to the expected threshold and results from statistical estimation of how a random distribution in overall level disables the strategy of using a loudness cue in consecutive comparisons (for details see ref. [1], pp. 19-21).

Experiments on profile analysis [1, 12-14] were devoted to investigating several signal parameters that might influence the discrimination process. These included a number of components and the frequency spacing between them, the spectral range occupied by stimuli components, the effects of masking of neighbouring components, in contrast to the importance of components remotely positioned on the frequency scale, the effect of pedestal (i.e., an initial increment in amplitude, relative to the amplitudes of other components). Although some studies were done for harmonic signals [15], most experiments on profile analysis were conducted with logarithmically spaced components in frequency what provided almost equal component distribution along the auditory filter bank. Within the scope of this study, this also corresponds to formants created in white or pink noise stimuli, using a 1/3-octave spectrum equalizer.

4. Is the formant detection/discrimination a real life realisation of the profile analysis task?

Laboratory experiments on profile analysis represent a quite theoretical construct of stimuli purposely designed for investigating the properties of the auditory system in its ability to analyse spectrally complex sounds. In this section an attempt is undertaken to directly compare formant detection, as it was done in the timbre solfege tasks, with some major results of profile analysis published by Green and Mason [1, 12], and Green and Kidd [14]. Actually, the two tasks will be compared. Firstly, the formant detection task in pink noise shown in Figure 1b will be compared with the profile analysis task involving the detection of an amplitude increment of a single sinusoid in a 21-component logarithmic complex [1, 12]. This comparison is shown in Figure 3. Secondly, the formant level discrimination task shown in Figure 2 will be compared with a profile analysis task involving detection of an increment in a 950-Hz component with pedestal, as this task corresponds to the discrimination of formant level [14]. This comparison is shown in Figure 4.

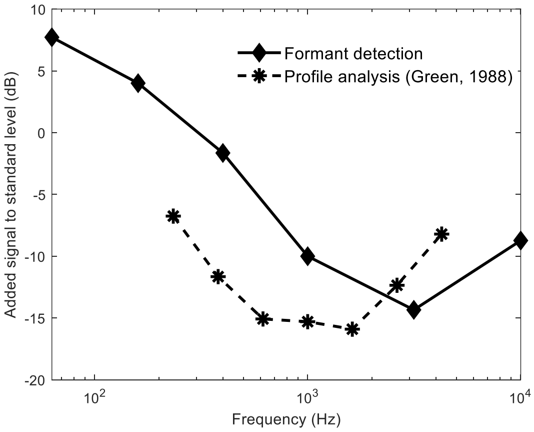

Figure 3. Comparison of a 1/3-octave formant detection task (solid line, diamonds) for white noise after Letowski and Rogala [8] with standard bowl of 21-compnent logarithmically spaced components in a range of 200–5000 Hz, in a profile analysis task after ref. [1] (dashed line, asterisks)

The solid line in Figure 3 presents the detection thresholds for 1/3-octave formants created in white noise, taken from Figure 2b and recalculated to the variable commonly used in profile analysis. In these studies, an increment ΔA in amplitude added in phase to the initial amplitude A of the component is a signal in threshold measurements. For example, a 0-dB signal level corresponds to $\Delta A = A$. Thus an addition results in doubling the signal component amplitude or a 6-dB increase in level. A signal level of –20 dB corresponds to $\Delta A = 0.1A$, causing the amplitude of the signal component to be increased by 10% or only 0.8 dB, relatively to the amplitudes of other spectral components. This is why thresholds in profile analysis tasks are usually negative in decibels, except for very large amplitude increments. The detection thresholds seen in Figure 1b, in a range from 10.7 dB at 63 Hz down to 1.5 dB at 3150 Hz, are plotted in new coordinates of ΔA level in Figure 3, in range from +7.7 dB down to -14.3 dB. The profile analysis data are taken from refs. [1, 12], and represent the well-established bowl for 21-component logarithmically spaced sinusoids (frequency ratio 1.1746 [12]). Unlike in the case of the data presented later in Figure 4, a direct comparison is possible, as both thresholds are defined for points located close to each other on the psychometric function: 75% correct for the formant detection task and 70.7% correct (adaptive procedure with a 2-down/1-up decision rule) in the profile bowl.

The variability of threshold with formant frequency, seen in Figure 3 for formant detection (solid line) is somewhat similar to the thresholds for amplitude change in the 21-component complex (dashed line). Both curves show a minimum of about –16 to –14 dB, although formant detection threshold at 3150 Hz, and profile analysis at 600–1700 Hz. The larger threshold increase for formant detection in low frequency range may be attributed to the energy decrease in narrowing the 1/3-octave band and thus an increase in masking by neighbouring bands. In contrast, log-spaced components in the auditory profile task produce approximately constant energy per critical band in the entire frequency range.

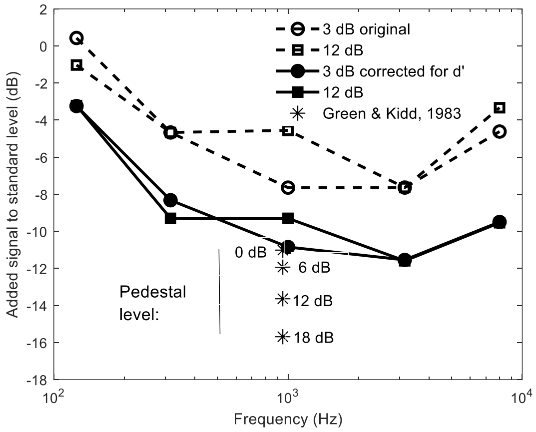

Figure 4. Comparison of 1/3-octave formant discrimination task (solid lines) and pedestal experiment by Green and Kidd [14] (asterisks). Dashed lines show experienced subjects' (M) original discrimination thresholds defined at 67% correct discrimination on the psychometric function. Solid lines show similar thresholds defined at 39% correct discrimination on a psychometric function corresponding to *d*'=0.74, the value identical to that of 2AFC task at 70.7% correct responses in the pedestal experiment

The analysis of discrimination thresholds shown in Figure 4 refers to the data presented in Figure 2, restricted to the group of experienced musicians. These data are compared with the pedestal experiment by Green and Kidd [14])] (asterisks in Figure 4). The dashed line in Figure 4 shows discrimination thresholds taken from Figure 2, expressed, as before, as 20log(Δ*A*/*A*). Due to differences in the experimental procedures, the values from experiments by Rogala and Śliwka [9] and Green and Kidd [14] had to be corrected to equalize the detectability index *d*' [16] before any comparison could be made. The thresholds in the pedestal experiment [14] correspond to *d*' = 0.74 as they were defined at 70.7% correct responses, as a result of an adaptive procedure with a two-down/one-up decision rule and a two-alternative forced choice paradigm of stimuli presentation. The formant experiment [9] used an *oddity* procedure for which *d*' = 0.74 occurs at 39% correct responses. Thus, using psychometric functions given in ref. [9], thresholds at 39% correct were determined and presented with solid lines in Figure 4.

The thresholds at 39% correct (solid lines) are on average by about 4 dB lower than those at 75% correct (dashed lines).

The corrected formant discrimination thresholds can be directly compared with the results of the pedestal experiment [14], of detecting a 950-Hz component in a 21-component complex (Figure 4, asterisks). Four pedestal levels, of 0, 6, 12 and 18 dB, were selected from ref. [14]. The average of thresholds for pedestals of 0 and 6 dB closely represent the initial formant level of 3 dB whereas the 12-dB pedestal corresponds to the 12-dB initial format level, to which an increment in level is added. It should be noted that formant discrimination thresholds closely correspond to 0 and 6-dB pedestals but are higher than the thresholds for the 12-dB pedestal. In the formant data reported by [9], the formant level discrimination thresholds are similar values for the 3- and 12-dB formants which was earlier interpreted as a finding indicating, that Weber's law holds for formant amplitude discrimination, which is not the case for the pedestal data. Overall, the data discussed in this paper suggest that the subjects concentrate in a formant discrimination task on the single frequency channel in which the change occurs and use sequential comparison, not simultaneous across-channel profile analysis. Such a listening strategy was very likely as the roving of signal level was not used in the experiments on formant discrimination.

5. Summary

This study was an attempt to compare the data on formant detection and discrimination obtained within the research studies concerned with the *Timbre Solfege* training program, developed by a team of researchers headed by Professor Andrzej Rakowski at the Musical Acoustics Laboratory, with a series of experiments within a research line called auditory profile, conducted by Professor David. M. Green and his associates. An added value of this comparison lies in the fact that the studies compared were, to a large extent, complementary in their research scope: the experiments with formants were conducted with close reference to auditory tasks commonly encountered in the praxis sound recording whereas the experiments on profile analysis used carefully designed stimuli aimed at studying specific properties of signal processing in the auditory system.

The major finding of this short study is that formant detection is based on inter-channel comparisons made independently for each of the stimuli compared in a trial. In contrast, formant discrimination consists in single channel comparison of amplitude at the formant frequency to detect changes in the formant amplitude and disregarding, as much as possible, the rest of the sound spectrum. This conclusion comes from the finding that the discrimination threshold for formant amplitude follows Weber's law which is not the case in auditory tasks based on profile analysis, involving across-channel comparisons. To verify these preliminary conclusions further experiments should be conducted at the Chopin University of Music with the use of some elements of profile analysis methodology.

Acknowledgments

Work supported by the grant 504/04064/1034/40.00 from the Warsaw University of Technology.

References

1. D. M. Green, *Profile analysis*, Oxford University Press, New York – Oxford 1988.
2. T. Letowski, A. Miśkiewicz, Timbre Solfege: *A course in perceptual analysis of sound*, In: *Signal Processing in Sound Engineering*, J. Adamczyk (Ed.), Warszawa: IPPT-PAN (2013) 83 – 96.
3. A. Rakowski, T. Łetowski, K. Szlifirski, B. Okoń-Makowska, *Developing sensitivity to the timbre of sound at the Department of Sound Engineering*, Academy of Music in Warsaw (in Polish), in: *Studia z Teorii Przekazu Dźwięku*, Wydawnictwa Radia i Telewizji, Warsaw, Poland (1982) 183 – 195.
4. T. Letowski, *Development of technical listening skills: Timbre solfeggio*, J. Aud. Eng. Soc., **33** (1985) 240 – 244.
5. A. Miśkiewicz, Timbre solfege: *A course in technical listening for sound engineers*, J. Aud. Eng. Soc., **40** (1992) 621 – 625.
6. T. Letowski, A. Miskiewicz, Developing of technical listening skills for sound quality assessment, Proceedings of *Inter-Noise* '95, Newport Beach, FL (1995) 917 – 920.
7. T. Rościszewska, A. Miśkiewicz, *Timbre Solfege: Development of auditory cues for the identification of spectral characteristics of sound*, Audio Engineering Society 138th Convention, May 7-10, Warsaw (2016).
8. T. Letowski, T. Rogala, *Formant perception: single formant*. In: *Sztuka słuchania* (*The Art of Listening*), Chopin University of Music, Warszawa (2015) 45 – 63.
9. T. Rogala, *Discrimination of formant frequency in pink noise*. Audio Engineering Society 140th Convention, June 4-7, Paris (2016) Paper #9583.
10. T. Rogala, *Pink noise bandwidth discrimination*, Audio Engineering Society 142th Convention, May 20-23, Berlin (2016).
11. T. Rogala, P. Śliwka. *Discrimination of formant amplitude in noise*, Audio Engineering Society 138th Convention, May 7-10, Warsaw (2016) Paper #9282.
12. D. M. Green, Ch. R. Mason, *Auditory profile analysis: frequency, phase, and Weber's law*, J. Acoust. Soc. Am., **77** (1985) 1155 – 1161.
13. D. M. Green, Ch. R. Mason, G. Kidd Jr., *Profile analysis: critical bands and duration*, J. Acoust. Soc. Am., **75** (1984) 1163 – 1167.
14. D. M. Green, G. Kidd Jr., *Further studies of auditory profile analysis*, J. Acoust. Soc. Am., 73 (1983) 1260 – 1265.
15. J. Żera, Z. A. Onsan, Q. T. Nguyen, D. M. Green, *Auditory profile analysis of harmonic signals: critical bands and duration*, J. Acoust. Soc. Am., **93** (1993) 3431 – 3441.
16. N. A. Macmillan, C. D. Creelman, *Detection theory: a user's guide*, Cambridge University Press, Cambridge 1991.