THE APPLICATION OF ILLUSIONS AND PSYCHOACOUSTICS TO SMALL LOUDSPEAKER CONFIGURATIONS

RONALD M. AARTS

Philips Research Europe, HTC 36 (WO 02) Eindhoven, The Netherlands

An overview of some auditory illusions is given, two of which will be considered in more detail for the application of small loudspeaker configurations. The requirements for a good sound reproduction system generally conflict with those of consumer products regarding both size and price. A possible solution lies in enhancing listener perception and reproduction of sound by exploiting a combination of psychoacoustics, loudspeaker configurations and digital signal processing. The first example is based on the missing fundamental concept, the second on the combination of frequency mapping and a special driver.

INTRODUCTION

A brief overview of some auditory illusions is given which serves merely as a 'catalogue', rather than a lengthy discussion. A related topic to auditory illusions is the interaction between different sensory modalities, e.g. sound and vision, a famous example is the McGurk effect ('Hearing lips and seeing voices') [1]. An auditory-visual overview is given in [2], a more general multisensory product perception in [3], and on spatial orientation in [4]. The influence of video quality on perceived audio quality is discussed in [5].
Two illusion based applications will be considered in more detail, in particular for the application of small loudspeaker configurations.

In many sound reproduction applications it is not possible to use large loudspeakers, due to size and/or cost constraints. Typical applications are portable audio, multimedia, TV and public address systems, to name just a few, see [6]. Hence the devices are often small in size, and therefore the transducers are inherently small as well. Needless to say, the competitive market also dictates these products achieve the highest possible audio quality. The most well-known characteristic of small loudspeakers is a poor low-frequency (bass) response. In practice this means that a significant portion of the audio signal may not be reproduced (sufficiently) by the loudspeaker. For loudspeakers used in such applications reproduction below 100 Hz is usually negligible, whereas in some applications of even smaller size this lower limit can easily be as high as several hundred hertz. The bass portion of an audio signal contributes significantly to the sound 'impact', and depending on the bass quality, the overall sound quality will shift up or down. Therefore a good low-frequency reproduction is essential.

ILLUSIONS

An illusion is a distortion of a sensory perception, revealing how the brain normally organizes and interprets sensory stimulation. While illusions distort reality, they are generally shared by most people [7]. Illusions can occur with each of the human senses, but visual illusions are the most well known and understood. However, we will focus on auditory illusions. First some overview is given on auditory illusions and then we will discuss how to utilize them. Two examples will be given; both have to do with bass reproduction. The first exploits the virtual pitch phenomenon the second the inaccuracy of human pitch detection at low frequencies.

A special issue of the J. Audio Eng. Soc. on auditory illusions is in [8]. Another overview or specific examples are given in [9,10]. Many auditory demonstrations including some illusions are on a CD [11] and a special CD on musical illusions and paradoxes is [12]. On the Internet one can find the most well-known demos [wl1-wl6].

Headphones vs. loudspeakers
It might seem trivial, but an important difference between headphones and loudspeaker listening is that the left channel sound radiated by a single left loudspeaker enters both ears, while with headphones listening, this sound reaches the left ear only. This has many consequences for sound localization; this is well treated in [13]. Another important difference is that with loudspeaker listening, room acoustics plays an important role, including the change of phase relations between the left and right signals or the phase between components within one signal.

Spatial illusions
Ordinary stereo reproduction is an example of a spatial illusion. Between and sometimes outside the loudspeakers, sources are audible, while there are no 'real' sources present, this is also covered in [13].

3D sound,
See 'Spatial illusions' above, and [14].

(Stereo) Base widening
Many schemes exist to give a virtual widening to loudspeaker base, see [15].

Elevation
The impression that the source is perceived to be located above the plane in which it is actually radiating is called elevation, see [16].

Missing fundamental
There is a vast amount of literature on this topic see e.g. [17]. If there are higher harmonics but the fundamental frequency is missing, we will still perceive the pitch of that missing fundamental, which in this case is also called *virtual pitch*. An application is the low frequency percept by small loudspeakers, see below in the section 'Virtual Pitch' and in [18, 19].

Doppler illusion
The Doppler illusion occurs if the sound source is moving with respect to the listener; see [20,21]. Strictly speaking this is not an illusion but a physical effect; nevertheless it is usually mentioned as an illusion.

Octave illusion
The octave illusion discovered by Diana Deutsch in 1973 [22, wl3] is an auditory illusion produced by simultaneously playing two sequences of two notes that are spaced an octave apart, high to low, and low to high, in separate stereo channels over headphones. People who are right-handed tend to hear the higher pitch as being in their right ear while the results are mixed for left-handed people, see also [23, 24].

Shepard's of pitch circularity
One of the most widely used auditory illusions is Shepard's [25,26] demonstration of pitch circularity, which has come to be known as the 'Shepard Scale' demonstration. The demonstration uses a cyclic set of complex tones, each composed of 10 partials separated by octave intervals. The frequencies of the partials are shifted upward in steps corresponding to a musical semitone (≈ 6 %). This creates the auditory illusion of a tone that continually ascends or descends in pitch, yet which ultimately seems to get no higher or lower.

Jean-Claude Risset subsequently created a version of the scale where the steps between each tone are continuous, and it is appropriately called the continuous Risset scale or Shepard-Risset glissando [wl6]. When done correctly, the tone appears to rise (or descend) continuously in pitch, yet return to its starting note. Risset has also created a similar effect with rhythm in which tempo seems to increase or decrease endlessly.

The glissando illusion
The glissando illusion was first reported and demonstrated by Diana Deutsch [27]. The auditory illusion is created when a sound with a fixed pitch, such as a synthesized oboe tone, is played together with a sine wave gliding up and down in pitch, and they are both switched back and forth between stereo loudspeakers. The effect is that the oboe is heard as switching between loudspeakers while the sine wave is heard as joined together seamlessly, and as moving around in space in accordance with its pitch motion. Right-handers often hear the glissando as travelling from left to right as its pitch glides from low to high, and then back from right to left as its pitch glides from high to low. Lefthanders often obtain different illusions.

Continuity
The illusory continuity of tones [28, 29] is the auditory illusion caused when a tone is interrupted for a short time (approximately 50 ms or less), during which a narrow band of noise is played. Whether the tone is of constant, rising or decreasing pitch, the ear perceives the tone as continuous if the 50 ms (or less) discontinuity is masked by noise.

McGurk effect
The McGurk effect [1] is a perceptual phenomenon which demonstrates an interaction between hearing and vision in speech perception. It suggests that speech perception is multimodal, that is, it involves information from more than one sensory modality. This effect may be experienced when a video of one phoneme's production is dubbed with a sound-

recording of a different phoneme being spoken. Often, the perceived phoneme is a third, intermediate phoneme. For example, a visual /ga/ combined with an audible /ba/ is often heard as /da/. Further research has shown that it can exist throughout whole sentences. The effect is very robust; that is, knowledge about it seems to have little effect on one's perception of it. This is different from certain optical illusions, which break down once one 'sees through' them.

Precedence effect
This effect states that it is the first acoustic information that arrives at a listener that determines the location of a sound [13].

Clifton effect
The effect, named after Rachael Clifton [30], introduces a change to the classical precedence demonstration. Half way into the click train, the source and echo clicks are reversed. That is, after several click-pair presentations as for the precedence effect (12 ms delay for the echo), the loudspeaker that had delivered the first or source click now delivers the delayed or echo click and vice a versa for the loudspeaker that delivered the echo click (it now delivers the first or source click). In the Clifton effect, an unexpected percept occurs immediately after the switch.

Franssen effect
The Franssen effect [13] named after Nico Franssen is a strong auditory illusion demonstrating the power of the first arriving information in establishing the location of a sound source. A tone is turned on abruptly at one loudspeaker and is then turned off slowly (with a 100-ms linear offset ramp). As this sound is going off, the sound is turned on at the other loudspeaker with the same envelope (with a 100-ms linear onset ramp). In this demonstration, this tone is usually left on for 5 seconds at this loudspeaker.

Ventriloquist
Ventriloquism is the ancient art of making one's voice appear to come from elsewhere. We regularly experience the effect when watching television and movies, where the voices seem to emanate from the actors' lips rather than from the actual sound source.

Auditory saltation
Auditory saltation is a mislocalization phenomenon in which click trains presented successively at two discrete spatial locations appear to originate from a smoothly varying series of source locations spanning the true source locations [36,wl7].

Sound-induced illusory flashing
This is a sound induced visual illusion. The illusion is: when a single flash is accompanied with two beeps, the single flash is perceived as two flashes. The illusion is strongest when the flash is in the periphery but it also works in the fovea. This illusion was discovered by Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo [wl1].

APPLICATIONS
Virtual Pitch
Pitch is a subjective, psychophysical quantity. According to the American Standards Association pitch is "that attribute of an auditory sensation in terms of which sounds may be ordered on a musical scale". For a pure tone, where the fundamental frequency corresponds to the frequency of the tone, the pitch is unambiguous and-if we neglect the influence of sound level on pitch-one can identify pitch with the frequency of the pure tone.

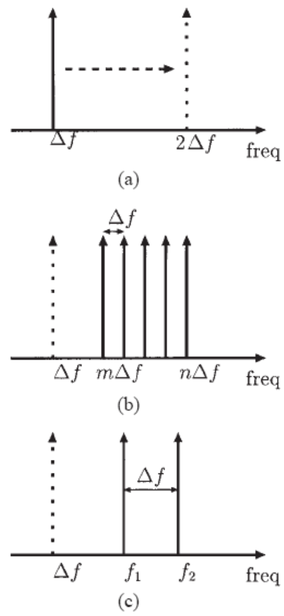

Fig. 1 Possible options for psychoacoustic bass enhancement. The dotted frequency component denotes the perceived pitch (but is not necessarily acoustically radiated). (a) Frequency doubling. (b) Residue pitch. (c) Difference tone

For a complex tone, consisting of more than one frequency, the situation is more complicated. Pitch should then be measured by psychophysical experiments.
A pitch that is produced by a set of frequency components, see Fig 1-b, rather than by a single sinusoid, is called a *residue*. In Fig.1-b the fundamental

frequency is missing, yet will still be perceived as a residue pitch, which in this case is also called *virtual pitch*. The psychoacoustic phenomenon responsible for this effect is called the 'missing fundamental' effect. There is long history of investigations into pitch perception, also regarding virtual pitch. Famous are the experiments of Seebeck in 1843, and the controversy of him with Ohm; see Plomp [17] for a historical review. There is a vast amount of literature on this topic, see [18,19] for some more references. When the frequency of a pure tone decreases to very low values, say less than 100 Hz, the pitch becomes more difficult to determine.

Figure 2 presents the general processing scheme that we propose for psychoacoustic bass enhancement. As the system is 'merely'' based on a psychoacoustic model of pitch perception, and uses loudspeaker characteristics in a very general sense (it is only assumed that reproducing lower frequencies is less efficient than reproducing higher frequencies), the method can be employed for any kind and/or size of loudspeaker.

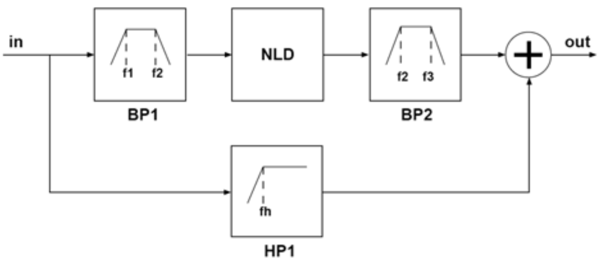

Fig. 2 Signal processing for psychoacoustic bass enhancement. The input signal is summed and filtered to obtain the bass portion. Then harmonics are created by the non-linear device (NLD) and added to left and right output signals. In the direct path a high-pass filter is implemented, from [19].

The non-linear device, or harmonics generator, 'shifts' signal components in a low frequency range to a higher frequency range. The pitch of the input signal is preserved, because the components in the higher frequency range are harmonics of the original components. The preservation of the original low pitch is due to the virtual pitch (or difference tone effect) of the harmonics signal. Because this element is a non-linear device, any single output component depends on all input components. Moreover, at the output, frequency components will be generated, which are not present at the input. This is a desired effect, since this is how the harmonics are obtained that create the illusion.

Small loudspeakers with high output

In the past much effort has been spent regarding the efficiency of loudspeakers, see e.g. [31]. However, focus in that work was on designing systems with a reasonably broad bandwidth, while for the present application we are aiming for the highest efficiency, which implies a very small bandwidth, and hence a high quality factor [32]. In [33] special resonant loudspeakers which have a high voltage sensitivity but require a normal cone displacement were discussed. In [34] a system which requires only a small cone displacement by using a tube as a resonator was introduced. Both systems have a similar way of working which is discussed below. For both of these special loudspeakers (see Fig. 3), a practically relevant optimality criterion, involving the loudspeaker and pipe parameters was discussed [34]. Common to both systems the impedance at the working frequency is ideally equal to exactly twice the DC impedance of the drivers.

Fig. 3 Picture of the prototype driver (MM3c) with a ten Euro cents coin. At the position where a normal loudspeaker has its heavy and expensive magnet, the prototype driver has an almost empty cavity; only a small moving magnet is necessary which is shown in the right corner.

The desired characteristics are obtained at the expense of a somewhat decreased sound quality and the requirement of some additional electronics. It is discussed how such a special loudspeaker can be made. It appears to be very cost-efficient and low-weight, and has a high degree in form factor freedom of its cabinet. An example of such a design is described and the performance of a working prototype was presented [33,34]. Due to the typical high and narrow peak in the frequency response, the normal operating range of the driver decreases considerably. This makes the driver unsuitable for normal use. To overcome this, a second measure is applied: The low-frequency content of the music signal, say 20—120 Hz, non-linear processing essentially compresses the bandwidth of a 20 to 120 Hz 2.5-octave bass signal down to a much narrower span, which is centred at the SPL peak of the system. This can be done with a set-up as depicted in Fig. 4 and will be discussed below. The band-pass filter takes the band of interest, typically 20—120Hz, and the envelope detector determines the envelope *m(t)* of this signal. Then *m(t)* is multiplied with a sinusoid of fixed amplitude and fixed frequency f_{work}. The result is that the coarse structure *m(t)* (the envelope) of the music signal after the compression or 'mapping' is the same as before the mapping. Only the fine structure has been changed to a sinusoid of fixed frequency f_{work} which coincides with the peak in the SPL response.

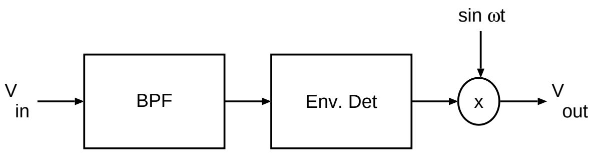

Fig. 4 Set-up of the mapping scheme. The boxes labelled 'BPF' and 'Env. Det.' are a band-pass filter and envelope detector, respectively. The sinusoid has a fixed frequency equal to f_{work}. The signal V_{out} is fed to the power amplifier and finally to the driver, from [18,33].

Many informal listening tests and demonstrations confirmed that the decrease in sound quality appears to be modest—apparently because the auditory system is less sensitive at low frequencies. Also, the other parts of the audio spectrum have a distracting influence on this mapping effect, which has been confirmed during formal listening tests [35], where the detectability of mistuned fundamental frequencies was determined for a variety of realistic complex signals.

CONCLUSIONS

We have discussed various auditory illusions. Two of them are considered in more detail including their applications. In the first example psycho-acoustic concepts are used to increase the bass perception. In the second example we replace the original low-frequency content with another signal adapted to a special efficient loudspeaker. This leads to an improved listening experience. This new loudspeaker together with some additional electronics yields a low-cost, lightweight, compact, sensitive, and efficient loudspeaker system that is very suitable for low-frequency sound reproduction. These advancements provide a large increase in form-factor design freedom.

REFERENCES

[1] H. McGurk and J. MacDonald, Hearing lips and seeing voices, Nature, December 1976, 264(23/30), pp. 746—748.

[2] A. Kohlrausch and S. van de Par, Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications, Proceedings of SPIE conference: Human Vision and Electronic Imaging, pp. 34—44, ed. T. N. Pappas and B. E. Rogowitz, 12 Apr. 1999.

[3] C. Spence and M. Zampini, Auditory contributions to multisensory product perception, Acta Acustica united with Acustica, 92, pp. 1009—1025, 2006.

[4] I.P. Howard and W.B. Templeton, Human spatial orientation, J. Wiley, 1966.

[5] J.G. Beerends and E. de Caluwe, The influence of video quality on perceived audio quality and vice versa, J. Audio Eng. Soc., 47(5), pp. 355—362, May 1999.

[6] R.M. Aarts, Methods to improve the sound reproduction of limited size systems, IEEE ICCE-07, Las Vegas, Jan. 10—14 2007.

[7] R.L. Solso, Cognitive psychology (6th ed.). Boston: Allyn and Bacon, ISBN 0-205-30937-2, 2001.

[8] J. Audio Eng. Soc., Auditory Illusions, 31(9) September 1983.

[9] D. Deutsch, Musical illusions, Scientific American, 233(4), pp.92—104, October 1975.
D. Deutsch, Two-channel listening to musical scales, J. Ac. Soc. Am., 57, pp. 1156—1160, 1975.

[10] R.M. Warren, Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception, J. Audio Eng. Soc., 31, pp. 623—629, September 1983.

[11] A.J.M. Houtsma, T.D. Rossing, and W.M. Wagenaars, Auditory Demonstrations, IPO through ASA, 1987.

[12] D. Deutsch, Musical Illusions and Paradoxes. Philomel Records. OCLC 36640949, 1995.

[13] J. Blauert, Spatial hearing Revised Edition: The Psychophysics of Human Sound Localization, The MIT Press, 1996.

[14] D.R. Begault, 3-D sound for virtual reality and multimedia, AP Professional, 1994.

[15] R.M. Aarts, Phantom sources applied to stereo-base widening, J. Audio Eng. Soc., 48(3), pp. 181—189, March 2000.

[16] P.J. Bloom, Creating source elevation illusions by spectral Manipulation, J. Audio Eng. Soc. 25(9), pp. 560—-565, Sept. 1977.
[17] R. Plomp, Experiments on tone perception, University Utrecht, 1966.

[18] E. Larsen and R.M. Aarts, Audio Bandwidth extension. Application of Psychoacoustics, Signal Processing and Loudspeaker Design, J. Wiley, ISBN 0470 85864 8, September 2004.

[19] E. Larsen and R.M. Aarts, Reproducing low-pitched signals through small loudspeakers, J. Audio Eng. Soc., 50(3), pp. 147—164, March 2002.

[20] J.G. Neuhoff and M.K. McBeath, The Doppler illusion: The influence of Dynamic intensity change on perceived pitch, Journal of experimental psychology: Human Perception and Performance, 22(4), pp. 970—985, 1996.

[21] J.G. Neuhoff and M.K. McBeath, Overcoming naive mental models in explaining the Doppler shift: An illusion creates confusion, American Journal of Psychology, 65(7), pp. 618—-621, July 1997.

[22] D. Deutsch, An auditory illusion, Nature, 251, pp. 307—309, September 1974.

[23] D. Deutsch, Auditory illusions, handedness and the spatial environment, J. Audio Eng. Soc., 31(9), pp. 607—618, September 1983.

[24] D. Deutsch, Lateralization and sequential relationships in the octave illusion, J. Ac. Soc. Am., 83(1), pp. 365—369, January 1988.

[25] R.N. Shepard, J. Audio Eng. Soc., Demonstrations of Circular Components of Pitch 31(9) pp. 641—649, September 1983.

[26] R. N. Shepard, Circularity in Judgments of Relative Pitch. J. Ac. Soc. Am., 36 (12), pp. 2346—2353, December 1964.

[27] D. Deutsch, K. Hamaoui, and T. Henthorn, The Glissando Illusion: A Spatial Illusory Contour in Hearing. Journal of the Acoustical Society of America, 117, p. 2476, 2005.

[28] R.M. Warren, J.M. Wrightson, and J. Puretz, Illusory continuity of tonal and infratonal periodic sounds, J. Ac. Soc. of Am. 84(4), pp. 1338—1342, October 1988.

[29] R. Plomp, demo CD How we hear. On the tone that makes the music (In Dutch: Hoe wij horen. Over de toon die de muziek maakt), 1998.

[30] R.K. Clifton, Breakdown of echo suppression in the precedence effect. J. Ac. Soc. Am., 82(5), 1834—1835, 1987.

[31] J. Vanderkooy, P.M. Boers, and R.M. Aarts, Direct-Radiator Loudspeaker Systems with High Bl, J. Audio Eng. Soc., 51(7/8), pp. 625—634, July/August 2003.

[32] R.M. Aarts, Optimally sensitive and efficient compact loudspeakers, J. Ac. Soc. Am., 119(2), pp. 890—896, Feb. 2006.

[33] R.M. Aarts, High-Efficiency Low-Bl Loudspeakers, J. Audio Eng. Soc., 53 (7), pp. 579—592, July/August 2005.

[34] R.M. Aarts, J.A.M. Nieuwendijk, and Okke Ouweltjes, Efficient Resonant Loudspeakers with Large Form-Factor Design Freedom. J. Audio Eng. Soc., 54(10), pp. 940—953, October 2006.

[35] N. le Goff, R.M. Aarts, and A.G. Kohlrausch, Thresholds for hearing mistuning of the fundamental component in a complex sound, Proceedings of the 18th International Congress on Acoustics (ICA2004), Paper Mo.P3.21, p. I-865, Kyoto, Japan, 2004.

[36] D. I. Shore, S. E. Hall, and R. M. Klein, Auditory saltation: A new measure for an old illusion, J. Ac. Soc. Am., 103(6), pp. 3730—3733, June 1998.

Web links

[wl1]
http://shamslab.psych.ucla.edu/demos/ referenced Feb. 20 2007.

[wl2]
http://www.cs.ubc.ca/nest/imager/contributions/flinn/Illusions/TT/tt.html referenced Feb. 22 2007.

[wl3]
http://psy.ucsd.edu/~ddeutsch/psychology/deutsch_research1.html referenced Feb. 22 2007.

[wl4]
http://www.kyushu-id.ac.jp/~ynhome/ENG/Demo/illusions.html referenced Feb. 22 2007.

[wl5]
http://littleshop.physics.colostate.edu/onlineexperiments/Auditory_Illusion.html referenced Feb. 22 2007.

[wl6]
http://www.exploratorium.edu/exhibits/highest_note/ex.about.fr.html referenced Feb. 22 2007.

[wl7]
http://www.utdallas.edu/~kilgard/Naturecorrespondence.html referenced Feb. 22 2007.