The effect of musical aptitude on the integration of audiovisual speech and non-speech signals in children

Kaisa Tiippana[1], *Kaupo Viitanen*[1], *Riia Kivimäki*[2]

[1]Department of Cognitive and Neuropsychology, Institute of Behavioural Sciences, P.O. Box 9, 00014 University of Helsinki, Finland
[2]Department of Music, University of Jyväskylä, Finland
kaisa.tiippana@helsinki.fi, kaupo.viitanen@helsinki.fi, kivimaki.riia@gmail.com

Abstract

Multisensory integration was assessed using two audiovisual illusions. In the McGurk effect, auditory speech perception is altered by incongruent visual speech. In the Shams illusion, the number of seen flashes is altered by an incongruent number of heard beeps. The illusions were tested in 10-year-old children, whose musical aptitude was also assessed. The strength of the McGurk effect was not linked to musical aptitude. However, children with high musical aptitude scores had a weaker Shams illusion and a narrower temporal window of integration, suggesting that they integrate non-speech information more selectively than children with low musical aptitude. These findings imply that musical aptitude influences multisensory integration selectively for rapid non-speech events.

Index Terms: audiovisual speech, McGurk effect, Shams illusion, multisensory integration, musical aptitude.

1. Introduction

The processing of audiovisual information differs between adult musicians and non-musicians. Auditory brainstem responses are larger and faster in musicians for audiovisual speech and musical stimuli [1]. The temporal window of audiovisual integration is narrower in musicians for musical but not for speech stimuli [2]. These findings suggest that musical expertise modulates audiovisual interactions.

The relationship between musical ability and audiovisual perceptual processing has not been studied previously in children. In the current study, we investigated whether musical aptitude in childhood is related to multisensory integration. It is possible that it is not only musical training that affects audiovisual perception and processing, but also musical aptitude, which can be considered a more innate trait [3].

There is evidence that there are links between linguistic and musical abilities. For example, school-aged children with advanced second language pronunciation skills have higher musical aptitude, as determined by Seashore's Test of Musical Talents [4], than children with a lower level of musical aptitude [5]. Also, musical aptitude correlates with phonological awareness in pre-school children [6]. It is thus possible that musical aptitude is related to speech perception also in the audiovisual domain.

We investigated whether musical aptitude influences audiovisual integration in school-aged children. Integration of audiovisual speech was assessed using the McGurk effect [7], where incongruent visual speech changes the auditory perception of a syllable, e.g. A[ba] presented with V[da] is heard as [da]. Sometimes this kind of a stimulus is heard as [g], producing a visually dominant auditory percept. Here we refer to the McGurk effect as an auditory percept differing from that produced by the acoustic signal when presented alone, be it a fusion or a visually dominant percept (see [8] for a discussion of this definition). The strength of the McGurk effect reflects the strength of audiovisual integration. That is, the fewer responses according to the acoustic stimulus are given, the stronger the visual influence and thus the McGurk effect, and the stronger the integration. If musical aptitude enhances audiovisual processing of speech, the McGurk effect should be stronger for children with high musical aptitude. On the other hand, if musical aptitude increases the weighting of auditory information in audiovisual integration, the McGurk effect might be weaker in children with high musical aptitude.

We also used a non-speech illusion where one quick flash is seen as two when accompanied by two rapid beeps (the Shams illusion [9]) to study whether the effect of musical aptitude is similar for the integration of both speech and non-speech stimuli. The Shams illusion is a non-speech phenomenon, which is not a direct control for the McGurk effect. It was chosen because it is a robust illusion like the McGurk effect. If there were only one integration mechanism, both illusions should behave similarly with respect to musical aptitude. If they do not behave similarly, the differences in integration may arise due to several reasons, for example because the Shams illusion is based on low-level simple non-speech stimuli requiring accurate temporal discrimination, meanwhile the McGurk effect is based on high-level complex speech stimuli requiring accurate spectral discrimination.

If musical aptitude increases the weighting of auditory information in audiovisual integration, the Shams illusion should be stronger in children with higher musical aptitude, indicating that audition influences perception more. On the other hand, since it has been shown that adult musicians have a narrower audiovisual temporal integration window for non-speech stimuli than non-musicians [2], children with high musical aptitude may process temporal information more accurately, i.e. their temporal integration window may be narrower than that of children with lower musical aptitude. If so, the former group of children might be more sensitive to audiovisual asynchrony and thus show a weaker Shams illusion than the latter. This is because there is always a delay between the two beeps, and an increase in this stimulus-onset-asynchrony (SOA) beyond about 100 ms decreases the illusion in adults [8]. In order to study the temporal window of audiovisual integration in the Shams illusion in more detail, we used two SOAs (short: 97 ms and long: 145 ms), with the hypothesis that if the integration window is narrower for children with high musical aptitude, the Shams illusion would diminish more rapidly for them with an increasing SOA.

In sum, we studied whether musical aptitude influences multisensory integration, and whether the effect is similar for speech and non-speech signals.

2. Methods

2.1. Participants

School children (n=49; 31 girls) aged 8-11, mean 10 years with Finnish as their mother tongue, reporting normal vision and hearing, participated in the experiments. Within this low age range, 51% of the children had not had any musical training, and the 49% who did, had from one up to five years of musical training (e.g. playing a musical instrument).

2.2. Stimuli and procedure

2.2.1. McGurk experiment

The stimuli were video recordings of a female uttering meaningless words [eme] and [ene]. They were presented in audio-only (A), visual-only (V) and audiovisual (AV) conditions. The audiovisual condition comprised the two original congruent utterances, and the McGurk stimulus. The McGurk stimulus was A[eme]V[ene], i.e. the voice saying [eme] was dubbed onto a face uttering [ene]. In the case of integration, this stimulus is heard as [ene], as has been shown previously for nasals [8]. There were 40 presentations of the McGurk stimulus and 20 of all other stimuli, in randomized order. The participants' task was to indicate whether they heard (A and AV) or lip-read (V) [eme] or [ene]. The height of the face was 55 mm on the computer screen. The speech sounds were played at 60 dBA.

2.2.2. Shams experiment

The visual stimulus was a white flash (duration 16 ms, diameter 22 mm, luminance 160 cd/m^2) in the centre of the computer screen. The auditory stimulus was a sinusoidal tone (duration 7 ms, 3500 Hz, 70 dBA). The stimuli were presented either once or twice with a stimulus-onset-asynchrony (SOA) of 97 or 145 ms. The stimuli were presented in audio-only, visual-only and audiovisual conditions. The AV condition comprised one flash paired with a simultaneous beep, two simultaneous flashes and beeps and the Shams illusion stimulus. The Shams illusion stimulus was one flash presented with two beeps, so that the onset of the flash and the first beep coincided, and the second beep followed after either 97 or 145 ms. There were 40 presentations of the Shams illusion stimulus, single flashes, single beeps and AV stimuli with one flash and beep, and 20 of all other stimuli, in randomised order for each delay. These repetitions were chosen to roughly equalize the percepts of one and two stimuli. The participants' task was to indicate whether they saw one or two flashes (V and AV), and, in the audio-only condition, whether they heard one or two beeps.

2.2.3. Apparatus

The visual stimuli were presented on the screen of a laptop computer placed at about 45 cm from the participant. The auditory stimuli were played via high-quality headphones. The experiments were run with Presentation 14.3 software (Neurobehavioral Systems).

2.3. Musical aptitude test

Musical aptitude was assessed using a combined score from selected sub-tests of the Seashore Measures of Musical Talents (SMMT) [4] and Montreal Battery of Evaluation of Amusia (MBEA) [10]. The SMMT sub-tests of 'pitch', 'time' and 'tonal memory' were used. The MBEA sub-test 'scale' was used. These were combined into a joint measure of musical aptitude ranging from 0 to 100%.

3. Results

3.1. Musical aptitude

The children were divided into two groups according to the score on the musical aptitude test. The limit dividing the low musical aptitude (L) and high musical aptitude (H) groups was chosen to be the median score of 62%, so that the children scoring 63% or over were assigned into the H group. This gave roughly equally sized groups, resulting in n=27 in the L group and n=22 in the H group.

3.2. McGurk experiment

The McGurk effect was rather strong in general, as evidenced by the quite low percentage of correct responses according to the auditory component: 63% (SEM 5) M-responses averaged across participants. The low and high aptitude groups did not differ significantly in the strength of the McGurk effect [$F(1,48)=0.62$, $p=0.43$], however with a trend towards a weaker effect in the low (66%, SEM 7) than high (58%, SEM 7) aptitude group (Fig. 1). The recognition of congruent audiovisual (L: 88%, SEM 1, H: 87%, SEM 1), auditory (L: 86%, SEM 1; H: 88%, SEM 1) and visual (L: 88%, SEM 2; H: 87%, SEM 2) stimuli did not differ between groups, either.

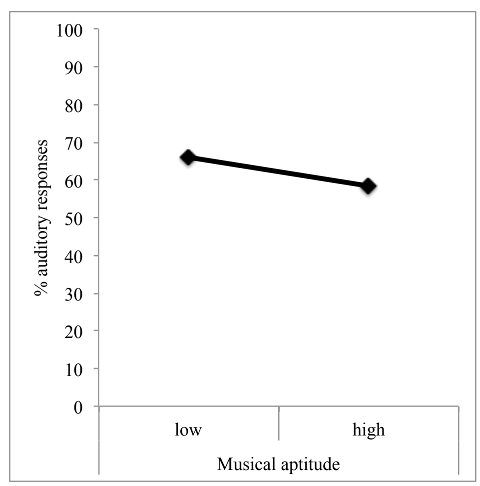

Figure 1: *McGurk experiment. The percentages of correct M-responses according to the auditory component for the McGurk stimulus A[m]V[n] for the groups with low or high musical aptitude. The lower the percentage, the stronger the McGurk effect. The groups did not differ significantly in the strength of the McGurk effect.*

3.3. Shams experiment

The Shams illusion was very strong in general, as evidenced by the low percentage of correct responses according to the visual component: 48% (SEM 4) of 1-responses averaged across conditions. This means that the children counted 2 flashes 52% of the time when presented with one flash and two beeps. The strength of the illusion was modulated by musical aptitude and SOA (short 97, long 145 ms), as shown in Figure 2. The Shams illusion was stronger for the shorter than longer SOA [$F(1,47)=30.08$, $p<0.001$], and approached significance for being stronger for the low than high aptitude group [$F(1,47)=3.73$, $p=0.059$]. A significant interaction [$F(1,47)=5.498$, $p=0.023$] arose because the difference between the groups was larger at the longer than shorter SOA. Follow-up univariate ANOVAs showed that the groups differed significantly only at the long [$F(1,47)=5.544$, $p=0.023$] but not at the short [$F(1,47)= 1.755$, $p=0.192$] SOA. That is, the group of children with high musical aptitude had a weaker Shams illusion at the SOA of 145 than the low aptitude group.

The recognition of congruent audiovisual (L: 87%, SEM, 2;H: 92%, SEM 2), auditory (L: 91%, SEM 1; H: 96%, SEM 2) and visual (L: 84%, SEM 2; H: 89%, SEM 2) stimuli did not differ between groups.

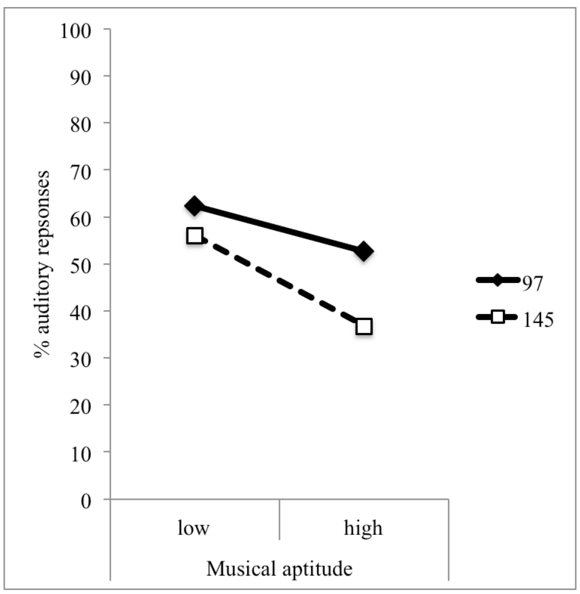

Figure 2: *Shams experiment. The percentages of incorrect 2-responses according to the auditory component for the Shams stimulus (A2beepsV1flash) for the groups with low or high musical aptitude at a short (97 ms) and long (145 ms) SOA, i.e. the delay between the two beeps. The higher the percentage, the stronger the Shams illusion. The illusion was significantly weaker in the high than low aptitude group at the long SOA.*

3.3.1. Contribution of musical training

The amount of musical training was unevenly distributed in the low and high musical aptitude groups. The children were split into two groups according to the amount of musical training: one group with one year or more of musical training, and the other group with no or less than a year of training. According to this division, 13 out of 22 in the H group had received musical training, while only 6 out of 27 in the L group had.

To test for the effect of musical aptitude without the effect of musical training, the results of the non-trained children (H: n=9; L: n=21) were re-analysed for the Shams experiment. The results were similar to the main analysis, with a stronger Shams illusion at the shorter SOA, and higher values for the low-aptitude group. The percentages of auditory responses were 59% (SEM 8) and 66% (SEM 5) for the H and L groups, respectively, at 97 SOA, and 45% (SEM 10) and 60% (SEM 6) for the H and L groups, respectively, at 145 SOA. However, only the main effect of SOA turned out to be significant in ANOVA [$F(1,28)=9.57$, $p=0.04$]. The lack of effect of low/high musical aptitude group implies that musical training, rather than musical aptitude, plays a major role in modulating the Shams illusion. Still, the numerical results were very similar in both analyses, and non-significant results in the latter may be caused by a lack of statistical power e.g. due to a low number of participants. It is also likely that the level of musical aptitude and musical training co-vary, and it is probably difficult to segregate them in a neat manner.

4. Discussion

Children in the high musical aptitude group had a weaker Shams illusion than children in the low musical aptitude group, meanwhile the groups did not differ significantly in the strength of the McGurk effect. The latter finding is surprising since links have been found between speech and musical abilities [5,6], so that an interaction could have been expected between audiovisual speech perception and musical aptitude. On the other hand, musical aptitude is largely an inborn trait [3], meanwhile multisensory integration generally develops with age and experience [11]. For example, the McGurk effect becomes stronger with age [7,12,13]. From this perspective, it is understandable that the aptitude to make musical discriminations is not linked to the integration of audiovisual speech. It is, however, possible that the capacity to integrate audiovisual information develops with experience also in the musical domain. Such integration is needed for example in playing an instrument by reading music [14,15]. Thus, audiovisual integration capacity might be linked to the amount of experience in playing a musical instrument, as has been shown in adults [1,2], and by the current analyses suggesting that musical training was a factor in the modulation of the Shams illusion. This issue should be addressed in more detail in the future.

We found evidence that the audiovisual temporal integration window was narrower for children with high rather than low musical aptitude, since an increasing temporal delay weakened the Shams illusion for the high aptitude group. This finding is in agreement with a previous finding of a narrow temporal window in adult musicians [2], and further suggests that the sharper temporal tuning is present early in life, even before extensive musical training.

A weak Shams illusion suggests poor audiovisual integration in the form of weak influence of audition on visual perception. The finding that the Shams illusion was weaker in children with high than low musical aptitude may first seem surprising, since musical aptitude could be expected to be linked to stronger audiovisual integration ability or auditory weighting in perception. On the other hand, musical aptitude in the current study refers to the ability to make temporal and pitch discriminations, so that the children with high aptitude are better at these discriminations. The Shams illusion is based on the discrimination of rapid auditory and visual events, i.e. on temporal discrimination. The better temporal discrimination ability that is linked with higher musical aptitude may be reflected in the Shams illusion as a better on-task performance, so that the high-aptitude children are more accurate in discriminating the visual stimuli as instructed, thus showing a weaker Shams illusion. The Shams illusion has been shown to be present and equally strong across age in children up to teen age [13], and actually stronger in children than adults [16]. It thus appears that the children with high musical aptitude display a more mature Shams illusion than children with low aptitude. Also, the trend of a stronger McGurk effect in these children implies that their multisensory integration is more mature than that of low-aptitude children.

The results of this study suggest that multisensory integration differs depending on the level of musical aptitude in school-aged children. This difference was found for audiovisual non-speech stimuli requiring rapid temporal discrimination, but not for audiovisual speech stimuli. It thus seems that musical aptitude affected a rather low-level integration process dealing with the temporal aspects of the signals, and not higher-level processes dealing with the semantic aspects of the signals that are required for speech categorization. However, it is very likely that learning and experience play an important role in the development of potential links between audiovisual integration and musical ability in a wider sense. Thus, the expected link between audiovisual speech perception and musical ability may be present at a later stage in the development. The interactions between multisensory integration and musical ability probably also depend on the musical and other experiences of the individual.

5. Conclusions

Integration of audiovisual speech was not affected by musical aptitude in school-aged children. In contrast, integration of rapid audiovisual non-speech signals was more selective in children with high musical aptitude than in children with low aptitude. Musical aptitude, apparently together with musical training, thus influenced multisensory temporal processing, without being reflected into the perception of audiovisual speech in childhood.

6. Acknowledgements

Kaisa Tiippana was supported by a three-year research grant from the University of Helsinki. Riia Kivimäki and Kaupo Viitanen were funded by the Centre of Excellence in Interdisciplinary Music Research, Academy of Finland.

7. References

[1] Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci USA, 104(40), 15894-15898.

[2] Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc Natl Acad Sci U S A, 108(51), E1441–E1450.

[3] Levitin, D. J. (2012). What does it mean to be musical? Neuron, 73(4), 633-637.

[4] Seashore, C. E., Lewis, D., & Saetveit, J. G. (1960). Manual of instruction and interpretations for the Seashore Measures of Musical Talents (2nd revision). New York: The Psychological Corporation.

[5] Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. Brain Res, 1194, 81-89.

[6] Peynircioglu, Z. F., Durgunoglu, A. Y., & Úney-Küsefoglu, B. (2002). Phonological awareness and musical aptitude. Journal of Research in Reading, 25(1), 68-80.

[7] McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264(5588), 746-748.

[8] Tiippana, K., Puharinen, H., Möttönen, R., & Sams, M. (2011). Sound location can influence audiovisual speech perception when spatial attention is manipulated. Seeing Perceiving, 24(1), 67-90.

[9] Shams, L., Kamitani, Y., & Shimojo, S. (2002). A visual illusion induced by sound. Cognitive Brain Research, 14, 147-152.

[10] Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. Annals of the New York Academy of Sciences, 999, 58-75.

[11] Ernst, M. O. (2008). Multisensory integration: a late bloomer. Curr Biol, 18(12), R519-521.

[12] Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. PLoS ONE, 2(1), e742.

[13] Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. Developmental Science, 11(2), 306-320.

[14] Schürmann, M., Raij, T., Fujiki, N. & Hari, R. (2002): Mind's ear in a musician: Where and when in the brain. NeuroImage 16, 434-440.

[15] Drai-Zerbib, V. r., Baccino, T., & Bigand, E. (2011). Sight-reading expertise: Cross-modality integration investigated using eye tracking. Psychology of Music, 40(2), 216-235.

[16] Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. Developmental Science, 14(5), 1089–1099.