languages

Supplementary Materials

1. Chi-square analyses

1.1. Results: Gender and singing behavior during childhood

Chi-square analysis has revealed that there was no association between gender and singing behavior during childhood $\chi^2(2) = 4.77$, $p = 0.92$.

1.2. Results: Musical status and singing behavior during childhood

Chi-square analysis has revealed that there was no association between the musical status (advanced musical, basic musical practitioners and non-musicians) and singing behavior during childhood $\chi^2(4) = 7.56$, $p = 0.11$.

2. Second statistical approach: the multi-item scale singing behavior during childhood

We also developed a multi-item scale concept which consisted of eight further questions on singing behavior during childhood as a re-test of our main questionnaire which defined the three distinct categories. These were "I sang very often (often)", "I sang a lot on an average amount (average)", and "I sang very seldom (seldom)". We, then, tested the internal consistency of our multi-item concept (Cronbach's alpha) and run correlational analysis between the multi-item scales concept and the variable which was created by the three response options (often, average, and seldom).

In a second step, we used the questions of the multi-item scale concept and run cluster analysis with the aim of creating meaningful groups. As a follow-up, we tested whether our groups differed in their mean values in all eight questions by running separate ANOVAs. Since cluster and the follow-up analyses provided meaningful three groups, we re-run the same analysis which has been provided in the method and results sections of the manuscript. First, we performed a MANOVA followed by discriminant analysis where the language perception, language pronunciation, singing capacity (singing complex) and music perception were the dependent variables.

2.1 Cronbach's α.

To assess the internal consistency of the questions, a reliability analysis was applied by means of performing Cronbach's α. The multi-item scale concept singing behavior during childhood shows high reliability, all Cronbach's $\alpha = .90$. Table S1 below illustrates the reliability analysis including the questions. Therefore, we coined a composite score which consisted of all eight questions.

Table S1. Translated questions of the multi-item scale concept singing behavior during childhood

Concept *Singing behavior during childhood*	
	Cronbach's α if item deleted
As a child I enthusiastically joined in with the singing at church and similar events whenever the possibility arose (Q1).	.887
As a child I was encouraged to sing by my caretakers and we sang together on a weekly basis even if there were no special events. (Q2).	.886
As a child I enjoyed singing in a choir, with friends, at Christmas, birthdays, or at similar occasions (Q3).	.884
As a child I sang very often since I wanted to become a musician or singer (Q4).	.896
As a child I used to sing whenever I could such as in the bathroom, in the car, when I played with friends (Q5).	.886
As a child I liked being a member of our school choir, or would have liked being a member or a school choir (Q6).	.899

As a child I enjoyed singing a song that had been played to me (e.g., in the radio) (Q7).	.876
As a child I used to sing more often than my friends (Q8).	.890

2.2 Correlations

In the next step, we correlated the multi-item concept with the singing behavior grouping variable (often, average, and seldom) which was indicated by the participants. Results revealed that the correlations between the composite score of the multi-item scales concept and the grouping variable singing behavior during childhood (often, average, and seldom) was rather high, $r = 0.88$, p (one-tailed) < 0.001.

2.3 Cluster analysis and follow-up ANOVAs

In order to test whether cluster groups can be differentiated based on the statements of the multi-item scale concepts, a hierarchical cluster analysis was applied, in which all eight questions were considered. We used Ward's methods (squared Euclidean distance), which resulted in three clusters of similar group sizes. Subsequent ANOVAs have revealed that all questions were significantly different in the three groups and showed the same patterns. Group 1 (N= 34) sang very often, group 2 (N = 29) sang an average amount and group 3 (N = 29) sang seldom. Table S2 shows the Welch's F-test, table S3 illustrates the descriptives and table S4 provides the results of the Games-Howell post-hoc analysis.

Table S2.Welch's F-test ANOVA singing concept.

Variables	**Welch's F**	p	ω
Q1	(2, 89) = 62.49	p < .001	0.76
Q2	(2,89) = 63.56	p < .001	0.76
Q3	(2, 89) = 73.13	p < .001	0.78
Q4	(1, 89) = 20.84	p < .001	0.55
Q5	(2, 89) = 30.96	p < .001	0.63
Q6	(2, 89) = 50.22	p < .001	0.72
Q7	(2, 89) = 178.44	p < .001	0.89
Q8	(1, 89) = 19.63	p < .001	0.54

Table S3. Descriptives of the questions Q1 -Q8.

Variables	**Group**	**Means and standard deviation**
Q1	often	M= 5.00; SD= 1.04
	average	M= 3.03; SD= 1.45
	seldom	M= 1.76; SD= 0.95
Q2	often	M= 4.94; SD= 1.18
	average	M= 3.38; SD= 1.50
	seldom	M= 1.55; SD= 0.78

Q3	often	M= 5.59; SD= 0.50
	average	M= 4.17; SD= 1.47
	seldom	M= 2.21; SD= 1.21
Q4	often	M= 3.82; SD= 1.83
	average	M= 2.38; SD= 1.68
	seldom	M= 1.41; SD= 0.57
Q5	often	M= 4.74; SD= 1.26
	average	M= 3.10; SD= 1.47
	seldom	M= 2.24; SD= 1.09
Q6	often	M= 5.06; SD= 1.41
	average	M= 3.24; SD= 1.99
	seldom	M= 1.38; SD= 0.62
Q7	often	M= 5.53; SD= 0.62
	average	M= 2.24; SD= 1.30
	seldom	M= 1.62; SD= 0.62
Q8	often	M= 4.47; SD= 1.40
	average	M= 3.17; SD= 1.37
	seldom	M= 2.41; SD= 1.57

Table S4. Games-Howell post-hoc analysis on the questions singing behavior during childhood.

Variables	**Post-hoc comparisons**	***t***	**df**	***p***	***r***
Q1	often vs. average	-6.69	89	< .001	0.58
	often vs. seldom	11.03	89	< .001	0.76
	average vs. seldom	-4.18	89	< .001	0.41
Q2	often vs. average	-5.20	89	< .001	0.48
	often vs. seldom	11.28	89	< .001	0.77
	average vs. seldom	-5.85	89	< .001	0.53

Q3	often vs. average	-5.06	89	< .001	0.47
	often vs. seldom	12.08	89	< .001	0.79
	average vs. seldom	-6.76	89	< .001	0.58
Q4	often vs. average	-3.82	89	< .001	0.38
	often vs. seldom	6.38	89	< .001	0.56
	average vs. seldom	-2.46	89	= .016	0.25
Q5	often vs. average	-5.03	89	< .001	0.47
	often vs. seldom	7.69	89	< .001	0.63
	average vs. seldom	0.88	89	= .012	0.09
Q6	often vs. average	-4.95	89	< .001	0.46
	often vs. seldom	10.02	89	< .001	0.73
	average vs. seldom	-4.88	89	< .001	0.46
Q7	often vs. average	-14.60	89	< .001	0.84
	often vs. seldom	17.36	89	< .001	0.88
	average vs. seldom	-2.65	89	= .009	0.27
Q8	often vs. average	-3.88	89	< .001	0.38
	often vs. seldom	6.15	89	< .001	0.55
	average vs. seldom	-2.18	89	= .032	0.23

2.4 MANOVA and discriminant analysis

We performed a MANOVA to understand whether our language perception, language pronunciation, singing ability, singing hours per week, the number of foreign languages and the music perception variables AMMA and tone frequency differ in their mean values. Using Pillai's trace there was a significant effect of singing frequency during childhood and language pronunciation and perception V = 0.441, $F(14,15) = 3.11$, $p < 0.001$.

As a follow-up, we also run discriminant analysis in order to compare the results of the childhood singing grouping variable with the cluster analysis-based grouping variable. The discriminant analysis- based on cluster analysis has provided almost identical results.

The first discriminant function explained 98.7% of the variance, canonical R^2= .40, whereas the second explained only 1.3%, canonical R^2= .001. In combination these discriminant functions significantly discriminated the groups, $\Lambda = 0.59$, $\chi^2(14) = 45.06$, $p < 0.001$, but removing the first function indicated that the second function did not significantly differentiate the three groups $\Lambda = 0.99$, $\chi^2(6) = .78$, $p = 0.993$. The correlations between the outcomes and the discriminant functions revealed that the loads onto the first function are rather high for the singing complex (r = .91) and the language pronunciation (r = .54). There were also correlations to the first function for singing hours per week (r = .41) and the language perception measures (r = -.14). Both were below the expected cutoff of .50 is used to decide which of the standardized discriminant coefficients are large, the singing complex and the language pronunciation highly discriminate the groups. This illustrates that the results of

the cluster-based approach provide almost the same findings as shown in the method section (see 3.3 in the manuscript).

3. Structure of the language perception measurement

The speech perception measurement consists of multiple items, and each item consists of a so called *Stimline* and a *Stimcompare*. A *Stimline* consists of a certain number (either 8, 10, or12) of *Stims*, where a *Stim* can be a word or short phrase in a foreign spoken by a native speaker of the respective languages. Each *Stim* of the *Stimline* is separated by a pause of 50ms. The words and phrases used for a *Stimline* were always taken from the same speaker for the respective language. A *Stimcompare* consists of 1, 2, or 3 Stims and is compared against the *Stimline* by the participant.

The participant listens to the *Stimline* first. 2 seconds after listening to the *Stimline* the participant listens to the *Stimcompare*. Then the participant has to tell whether the *Stimcompare* was/were included in the *Stimline* or not, and indicate how certain they are on a scale from 1 to 3 ("very sure", "rather sure", "can't tell"). After the test is complete, a score is calculated according to the number of correct answers given. A familiarisation task at the start of the test is not counted, and the items are presented to the participants in randomised order. Twelve of the *Stimcompare* are a match, six of the *Stimcompare* represent a partial match, while six of the *Stimcompare* do not match. The 3 types of *Stimcompares* (1, 2 or 3 Stims) are categorised into subsections (A, B and D). The letter C is omitted to avoid confusion with the meaning of "correct" for the answer given.

① **S-Id:** *Stimline*/*Stimcompare*-Identifier

② **SCS:** *Stimcompare* subsection
The subsection of the *Stimcompare*; within the *Stimline* this value is 0, within the *Stimcompares* it describes the subsection (A, B, D).

③ **SCNR:** *Stimcompare* number
Number of the subsection of the *Stimcompare*

④ **Pos:** Position of the Stim within the *Stimline*/*Stimcompare*

⑤ **Language:** Language of the *Stim*

⑥ **Syll:** Syllable length of the *Stim*

⑦ **Stim-No:** Recording number of the *Stim* with the same syllables

⑧ **Identifier:** Database identifier of the *Stim*

⑨ **f_c:** "false" or "correct"
Printing "c" or "f" in the *Stimcompare* section, see Ⓒ
If "false" the correct answer to the item is "not included in *Stimline*", if "correct" the correct answer to the item is "is included in *Stimline*".

⑩ **SL_Pos:** the position of the Stim within the *Stimline*.
For the *Stimline* section this number is the same as in ④ but for the *Stimcompare* it shows whether the *Stim* is included in the *Stimline* – the position number in the *Stimline* is printed – or if the *Stim* is not included in the *Stimline* – there is no number printed.

Ⓐ **The Stimline section**
e.g. "SL16" read: *Stimline* 16

Ⓑ **The Stimcompare section**
e.g. "SC-16-A1" read: *Stimcompare* to *Stimline* 16, subsection A, number 1

Ⓒ **false or correct**, see ⑨

The example provided below shows a sample of the Japanese perception task. The Stimline consists of 8 Stims and the Stimcompare consists of 2 Stims. These 2 Stims can either both included in the Stimline, one of them can be included in the Stimline, or none of them can be included in the Stimline. The participant is then asked whether the Stimcompare was fully included, and can answer either "yes" or "no". The answer "yes" is only correct if all 2 Stims of the Stimcompare were included in the Stimline, and the answer "no" for all other cases.

Aptitude test NV **Stimline 16**

① S-Id	② SCS	③ SCNR	④ Pos	⑤ Language	⑥ Syll	⑦ Stim-No	⑧ Identifier	⑨ f_c	⑩ SL-Pos
Ⓐ **SL16**									
	0	0	1	Japanese	5	2	JA-05-02		1
	0	0	2	Japanese	7	2	JA-07-02		2
	0	0	3	Japanese	7	1	JA-07-01		3
	0	0	4	Japanese	4	2	JA-04-02		4
	0	0	5	Japanese	9	1	JA-09-01		5
	0	0	6	Japanese	7	3	JA-07-03		6
	0	0	7	Japanese	4	3	JA-04-03		7
	0	0	8	Japanese	5	3	JA-05-03		8
Ⓑ **SC16-A1** **c**								Ⓒ	
	A	1	1	Japanese	5	2	JA-05-02	c	1
SC16-A2 **f**									
	A	2	1	Japanese	5	1	JA-05-01	f	
SC16-B1 **c**									
	B	1	1	Japanese	4	2	JA-04-02	c	4
	B	1	2	Japanese	7	3	JA-07-03	c	6
SC16-B2 **f**									
	B	2	1	Japanese	5	1	JA-05-01	f	
	B	2	2	Japanese	7	1	JA-07-01	f	3
SC16-B3 **c**									
	B	3	1	Japanese	5	2	JA-05-02	c	1
	B	3	2	Japanese	7	3	JA-07-03	c	6
SC16-D3 **c**									
	D	3	1	Japanese	4	2	JA-04-02	c	4
	D	3	2	Japanese	7	3	JA-07-03	c	6
	D	3	3	Japanese	5	2	JA-05-02	c	1
SC16-D4 **f**									
	D	4	1	Japanese	4	3	JA-04-03	f	7
	D	4	2	Japanese	5	1	JA-05-01	f	
	D	4	3	Japanese	7	1	JA-07-01	f	3