Supplemental Material S1. Summary of musician's advantage for speech-in-speech.

Study	Mean Age M = musician NM = nonmusician	Musician criteria	Paradigm(s)	Musician's Advantage?
Parbery-Clark, Skoe, Lam, & Kraus (2009)	M/NM: 23 ± 3 y (range: 19-31)	≥ 10 y training Started ≤ age 7 Practice ≥ 3xs/wk	QuickSIN Repeat sentences presented in 4 talker babble (composed of 3 female voices, 1 male voice), varying the signal to noise ratio (SNR).	**Yes**
Parbery-Clark, Strait, Anderson, Hittner, & Kraus (2011)	M: 55 ± 4.24 y NM: 54 ± 6.02 y	Started ≤ age 9 Practice ≥ 3xs/wk	WIN & QuickSIN Words and sentences in 4-talker babble	**Yes** (both tasks)
Zendel and Alain (2012)	M: 45.3 y (range: 19-91) NM: 49.3 y (range: 18-86)	Started ≤ age 16 ≥ 6 y lessons	QuickSIN Sentences in 4-talker babble	**Yes** (for older musicians)
Strait, Parbery- Clark, O'Connell, & Kraus, (2013)	M/NM: 3-5 y	Started training ≤ 12 months prior Weekly lessons Practice ≥ 4xs/wk	ABR /da/ Hear a syllable in quiet and/or in babble and record brainstem response via electroencephalogram (EEG). Presented in 6-talker babble (2 male voices, 2 females voices)	**Yes**
Ruggles, Freyman, & Oxenham (2014)	M: 21.8 y NM: 20.7 y	≥ 10 y training Started ≤ age 10 Practice ≥ 5 hr/wk	QuickSIN Sentences in 4-talker babble	**No difference**
Boebinger et al. (2015)	M/NM: 27.2 ± 6.9 y	≥ 10 y training Started ≤ age 7 Practice ≥ 3 xs/wk	BKB sentence targets Spoken by a female speaker. Presented with a male masker	**No difference**
Zendel et al. (2015)	M: 23.4 ± 4.3 y NM: 21.9 ± 2.6 y	≥ 10 y training Started ≤ age 15 Practice ≥ 10 hr/wk	CVC words Presented in 4-talker babble (15 dB SNR, 0 dB SNR)	**Yes** (only at 0 dB SNR)
Anaya et al. (2016)	M/NM: 20.72 ± 2.72 y	Started ≤ age 9 Enrolled in college music program	PRESTO Sentences presented in 6-talker babble	**No difference** (for composite speech-in-speech + speech-in-noise score)
Başkent & Gaudrain (2016)	M: 22.75 ± 2.43 y NM: 21.89 ± 1.97 y	≥ 10 y training Started ≤ age 7	Versfeld et al. (2000) sentences produced in 1-talker babble Masker created by concatenating random 1 s sequences of non-target sentences. Mean f0 and apparent vocal tract length were manipulated.	**Yes**
Clayton et al. (2016)	M: 22.5 ± 2.8 y NM: 20.47 ± 1.4 y	≥ 10 y training Practice ≥ 5 hr/wk Enrolled in college music program	Target and 1-talker masker sentences (Swaminathan et al., 2015). Targets presented at 0°, while makers were either also presented at the same spatial location or at ±15°. Target and maskers were recorded by different female talkers. Target sentences were cued by the call-sign 'Jane.'	**Yes** (when masker was spatially separated; no advantage when target and masker were at 0°)
Mandikal Vasuki, Mridula Sharma, Demuth, & Arciuli (2016)	M: 28 y (median) NM: 25 y (median)	≥ 10 y training Started ≤ age 9	LiSN-S test Repeat sentences produced by the same or a different talker at (0°)	**No difference**
Slater & Kraus (2016)	M: 25.4±5.7 y (Percussionists) 23.4±3.6 y (Vocalists) NM: 23.2±3.8 y	Active musicians ≥ 7 y	WIN & QuickSIN Words and sentences in 4-talker babble	**Yes** (QuickSIN for drummers only. No difference for WIN)
Deroche, Limb,	M: 21.9 ± 2.6 y	8 y training	Harvard/IEEE Sentence for target and	**No difference**

Chatterjee, & Gracco (2017)	NM: 25.1 ± 5.9 y	Started ≤ age 8	2-talker maskers spoken by the same male talker The masker had a fixed f0 (150 Hz), while the target f0 varied (Δf0 = 0, -2, -8 ST). Target and maskers were also presented in the same and different ears.	(no musician effect or interaction with f0 or ear)
Madsen, Whiteford, & Oxenham (2017)	M: 21.13 ± 2.47 y NM: 20.9 ± 2.70	10 y training Started ≤ age 7 Practice ≥ 5 hr/wk	Target (HINT) sentences with 1-talker interferer (IEEE sentences) Target and masker recorded by different male talkers. The masker average f0 was lower than the target by 0, 1, 2, 4, 8 ST (1/2 trials with normal intonation and 1/2 of trials monotone).	**No difference** (for both natural and monotone f0 conditions)
Morse-Fortier, Parrish, Baran, & Freyman (2017)	M: 20.1 y NM: 22.5 y	Daily practice Enrolled in college music program	Target sentences with 2-talker masker sentences (from Helfer, 1997) Monitored for words from a list. Target voice was a female talker. Maskers were 2 other female talkers.	**Yes**
Yeend et al. (2017)	M/NM: 45 y (range: 30-57)	Professional musicians	NAL-DCT Monologues presented in multitalker background speech (-7dB SNR) LiSN-S test Repeat sentences produced by a different talker at (± 90°)	**No difference** (both tasks)
Zendel, West, Belleville, & Peretz (2017)	Musical training group: 67.5 ± 4.2 y Control: 69.3 ± 5.7 y	All were nonmusicians (≤ 3 y musical training)	Monosyllabic words in multitalker babble Babble created by combining monologues spoken by 4 speakers at 15dB or 0 dB SNR	**Yes** (Musical training group showed more improvement)
Başkent et al. (2018)	M: 12.4 y (range = 11-13) NM: 12.3 y (range: 11-14)	≥ 5 y training Started ≤ age 7 Musical training in the last 3 y	Meaningful target sentences with a masker (concatenated partial sentences). Target spoken by a female talker. Masker either by the same female talker or a male talker (masker onset preceded target sentence).	**No difference**
Couth et al. (2021)	M: 18-26 y NM: 18-27 y	College/early-career musicians (either completing or graduated < 1 y prior)	CRM paradigm Target cued by 'Baron'; two maskers. Target and masker talkers were randomly selected from 2 male and 2 female talkers	**No difference**
Kaplan et al. (2021)	M: 27.13 y (range: 19-45) NM: 26.35 y (range 19-46)	≥ 10 y training Started ≤ age 7 Practicing ≥ 3 y prior to the study	Semantically neutral target sentences with 1-talker masker (meaningful sentences from Versfeld et al., 2000). Target and masker were recorded by 2 female talkers. Target-to-masker ratio (TMR): -3dB, -5 dB, -7dB, -9dB	**Yes**
Mussoi (2021)	M: 69.5 ± 4.5 y NM: 70.1 ± 3.6 y	≥ 5 y training Started ≤ age 10 Practice ≥ 3 hr/wk	QuickSIN Words in 4-talker babble	**No difference**

Supplemental Material S2. Calculations of semitone separation based on Kishon-Rabin et al. (2001).

Group	Relative difference limen (relDLF); $\Delta f/f_1$	Just noticeable difference (JND) relative to 100 Hz	Semitone difference from 100 Hz; *hqmisc* R package *f2st*($f_{2,}$, base = 100)
Musicians	$\Delta f/f_1 = 0.00907$	$\Delta f = 0.00907 : f_{100\ \mathrm{Hz}}$ $\Delta f = 0.907$ $f_2 = 100.907$	ΔST = 0.156
Nonmusicians	$\Delta f/f_1 = 0.01783$	$\Delta f = 0.01783 : f_{100\ \mathrm{Hz}}$ $\Delta f = 1.783$ $f_2 = 101.783$	ΔST = 0.306

Kishon-Rabin et al. (2001) found that musicians had a smaller relative difference limen (refDLF: $\Delta f/f_1$ =0.00907) than nonmusicians (relDLF: $\Delta f/f_1$ = 0.01783) in perceiving a difference in pure tones. We calculated what this difference limen would be relative to 100 Hz (Δf =relDLF: $f_{100\ \mathrm{Hz}}$). We then calculated the difference in semitones between the just-noticeable difference (JND) frequency (100.907 Hz for musicians, 101.783 for nonmusicians) and starting frequency (100 Hz) with the *hqmisc* R package: *f2st(100 Hz + Δf , base = 100 Hz)*.

Supplemental Material S3. **Sentence identification** (Experiment 1): Posterior means (Estimate), standard deviation of the posterior (Error), 95% credible intervals (Q2.5, Q97.5), and percent of posterior distribution above or below zero, for fixed effects. Effects whose credible intervals do not include zero, or those with 95% of their distribution on one side of 0 are in bold.

					% Distribution	
	Estimate	Error	Q2.5	Q97.5	< 0	> 0
Intercept	**-0.83**	**0.28**	**-1.38**	**-0.29**	**100**	**0**
Group (Musician)	0.04	0.04	-0.04	0.12	17	83
Age (YA)	**0.13**	**0.05**	**0.04**	**0.22**	**0**	**100**
F0	**0.27**	**0.04**	**0.19**	**0.34**	**0**	**100**
Block	**0.03**	**0.01**	**0**	**0.05**	**1**	**99**
SingleSentenceAcc	-0.04	0.05	-0.13	0.05	83	17
Group(Musician):Age(YA)	**0.1**	**0.04**	**0.02**	**0.19**	**1**	**99**
Group(Musician):F0	0.02	0.04	-0.06	0.09	34	66
Age(YA):F0	0.05	0.04	-0.02	0.13	8	92
Group(Musician):Age(YA):F0	**0.06**	**0.04**	**-0.01**	**0.14**	**5**	**95**
Num. observations = 9,792 ; *Num participants* = 51 ; *Num. sentences* = 16						

Supplemental Material S4. Confusion matrix for participants who did not reach 90% in single vowel identification (shown in percentages).

	observed				
expected	bat	bought	bet	beat	boot
/æ/	69.7%	5.3%	22.7%	2.3%	0%
/ɑ/	66.7%	31.1%	1.5%	0.8%	0%
/ɛ/	6.1%	6.1%	78%	6.8%	3%
/i/	0%	1.5%	15.9%	81.1%	1.5%
/u/	0.8%	22.7%	2.3%	0%	74.2%

Supplemental Material S5. Confusion matrix for YA nonmusicians who did reach 90% in single vowel identification (shown in percentages).

	observed				
expected	bat	bought	bet	beat	boot
/æ/	89.7%	0%	10.3%	0%	0%
/ɑ/	23.1%	76.9%	0%	0%	0%
/ɛ/	0%	1.3%	94.9%	2.6%	1.3%
/i/	0%	0%	2.6%	97.4%	0%
/u/	0%	1.3%	5.1%	0%	93.6%

Supplemental Material S6. Confusion matrix for YA musicians who did reach 90% in single vowel identification (shown in percentages).

	observed				
expected	bat	bought	bet	beat	boot
/æ/	97.9%	0%	2.1%	0%	0%
/ɑ/	11.5%	88.5%	0%	0%	0%
/ɛ/	0%	0%	100%	0%	0%
/i/	0%	0%	0%	100%	0%
/u/	0%	0%	0%	0%	100%

Supplemental Material S7. Confusion matrix for OA nonmusicians who did reach 90% in single v owel identification (shown in percentages).

	observed				
expected	bat	bought	bet	beat	boot
/æ/	94%	0%	6%	0%	0%
/ɑ/	10.7%	88.1%	1.2%	0%	0%
/ɛ/	1.2%	4.8%	91.7%	2.4%	0%
/i/	1.2%	0%	6%	92.9%	0%
/u/	0%	3.6%	1.2%	1.2%	94%

Supplemental Material S8. Confusion matrix for OA musicians who did reach 90% in single vowel identification (shown in percentages).

	observed				
expected	bat	bought	bet	beat	boot
/æ/	96.4%	0%	2.4%	1.2%	0%
/ɑ/	16.7%	83.3%	0%	0%	0%
/ɛ/	0%	0%	97.6%	0%	2.4%
/i/	1.2%	0%	0%	98.8%	0%
/u/	0%	0%	4.8%	0%	95.2%

Supplemental Material S9. Double vowel identification (Experiment 2): Posterior means (Estimate), standard deviation of the posterior (Error), 95% credible intervals (Q2.5, Q97.5), and percent of posterior distribution above or below zero, for fixed effects. Effects whose credible intervals do not include zero, or those with 95% of their distribution on one side of 0 are in bold.

					% Distribution	
	Estimate	Error	Q2.5	Q97.5	< 0	> 0
Intercept	**-1.19**	**0.21**	**-1.6**	**-0.77**	**100**	**0**
Group(Musician)	0.1	0.13	-0.16	0.36	22	78
F0	**0.45**	**0.04**	**0.37**	**0.52**	**0**	**100**
Age(YA)	-0.1	0.13	-0.36	0.16	78	22
Block	**0.09**	**0.01**	**0.07**	**0.11**	**0**	**100**
VowelDistance	**0.38**	**0.19**	**0**	**0.75**	**3**	**97**
JointSingleVowelAcc	**0.15**	**0.02**	**0.11**	**0.2**	**0**	**100**
Group(Musician):F0	0	0.04	-0.07	0.07	53	47
Group(Musician):Age(YA)	0.15	0.13	-0.1	0.41	12	88
F0:Age(YA)	**-0.09**	**0.04**	**-0.16**	**-0.02**	**99**	**1**
Group(Musician):VowelDistance	0.03	0.18	-0.33	0.4	43	57
Age(YA):VowelDistance	-0.04	0.18	-0.4	0.32	59	41
Group(Musician):F0:Age(YA)	0	0.04	-0.07	0.07	49	51
Group(Musician):Age(YA): VowelDistance	-0.05	0.18	-0.41	0.31	61	39

Num. observations =12,000; *Num. participants* = 50; *Num. vowel pairs* = 20