

This unedited manuscript has been submitted for publication in the Annals of the NYAS. This paper has not been copyedited.

The OPERA hypothesis: assumptions and clarifications

Journal:	<i>Annals of the New York Academy of Sciences</i>
Manuscript ID:	Draft
Manuscript Type:	Conference papers
Date Submitted by the Author:	n/a
Complete List of Authors:	Patel, Aniruddh; The Neurosciences Institute, Theoretical Neurobiology
Keywords:	musical training, speech perception, neural plasticity, neural encoding, pitch

SCHOLARONE™
Manuscripts

Patel, A.D. : Refining the OPERA hypothesis

The OPERA hypothesis: assumptions and clarifications

Aniruddh D. Patel

¹The Neurosciences Institute

Corresponding author: Aniruddh D. Patel, The Neurosciences Institute, 10640 John Jay

Hopkins Dr., San Diego, CA 92121, USA. Tel: 858-626-2085, Fax: 858-626-2099.

apatel@nsi.edu

Keywords: Musical training, speech perception, neural plasticity, neural encoding, pitch perception

Submitted to *Annals of the New York Academy of Sciences*

Oct 31, 2011

1
2
3 **Abstract:** Recent research suggests that musical training enhances the neural encoding of
4 speech. Why would musical training have this effect? The OPERA hypothesis proposes
5 an answer, based on the idea that musical training demands greater precision in certain
6 aspects of auditory processing than does ordinary speech perception. The current paper
7 presents two assumptions underlying this idea, as well as two clarifications, and suggests
8 directions for future research.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Recent neuroscientific research has revealed that musically trained individuals show enhanced neural encoding of speech sounds in the auditory brainstem relative to musically untrained individuals.¹ One possible source of this difference is innate differences in the structure and function of the auditory system between those who do vs. not pursue musical training. Another possible source is experience-dependent neural plasticity due to musical training. A role for plasticity is suggested by the repeated finding that the degree of enhancement correlates with the number of years of musical training. This plasticity may be driven by the extensive corticofugal pathways which originate in auditory cortex and project to subcortical auditory regions.²

Firm evidence that musical training causes enhancements in brainstem speech encoding awaits controlled studies in which individuals are randomly assigned to musical training vs. some other enjoyable artistic training (such as painting), and brainstem processing of speech is measured before and after training.^{3,4} Hopefully such research will be conducted soon. It seems very likely that such studies will demonstrate a causal role for musical training in enhancing the quality of brainstem speech sound encoding. This will be an important result because the quality of brainstem speech sound encoding is correlated with real-world language skills such as reading ability and hearing speech in noise.¹ This in turn suggests that musical training may be useful for improving these skills in normal individuals or in those who suffer from developmental problems with these skills.

1
2
3
4 If musical training causes enhancements in subcortical speech encoding, this
5
6 raises a fundamental question: *Why* would musical training have this effect? After all,
7
8 musical instruments (such as guitars or trumpets) sound very different from human voices.
9
10 Furthermore, spoken sentences are acoustically very complex, and many musical
11
12 instruments produce sound patterns that are simpler by comparison (e.g., consider the
13
14 acoustics of a flute melody vs. a speaking voice). Why would learning to produce and
15
16 perceive relatively simple acoustic patterns improve the brain's processing of complex
17
18 acoustic patterns?
19
20

21
22 One possible explanation is offered by the OPERA hypothesis.⁵ OPERA argues
23
24 that musical training enhances the neural encoding of speech when five conditions are
25
26 met: Overlap, Precision, Emotion, Repetition, and Attention. Together, these conditions
27
28 drive adaptive neural plasticity in auditory processing networks, leading them to function
29
30 with higher precision than needed for ordinary speech communication. Yet, since speech
31
32 shares these networks with music, speech processing benefits.
33
34

35
36 Of these five conditions, the primary novel idea is that musical training impacts
37
38 speech processing because it demands greater precision than speech perception in certain
39
40 aspects of auditory processing. This paper focuses on this issue of precision, discussing
41
42 two assumptions underlying OPERA which were not made explicit in the original
43
44 statement of the hypothesis, as well as two clarifications. The paper focuses on pitch
45
46 processing, an important part of both speech and music.⁶ The auditory system has neural
47
48 mechanisms for extracting pitch from complex sounds,⁷ and it is very likely that common
49
50 brainstem circuits are involved in pitch processing in speech and music (satisfying the
51
52 'Overlap' condition of OPERA).
53
54
55
56
57
58
59
60

1
2
3 Before turning to the assumptions and clarifications mentioned above, it is worth
4 reviewing the idea that musical training demands greater precision than speech
5
6 perception in terms of pitch processing. One way to approach this issue is to ask: “how
7
8 important is hearing fine pitch distinctions for speech comprehension vs. for musical
9
10 training?”
11
12
13

14
15 If a person cannot hear fine distinctions in pitch, the chances are that he or she
16
17 will still comprehend the basic meaning of spoken sentences. This is because speech has
18
19 many redundant cues to meaning. For example, if a listener doesn’t perceive this pitch
20
21 rise at the end of “Is it your birthday?”, he or she will still know the sentence is a question
22
23 by virtue of word choice (“Is it...”) and grammar. Indeed, it has recently been shown
24
25 that native speakers of Mandarin Chinese sentences can understand sentences with no
26
27 pitch variation (i.e., native listeners find monotone Mandarin sentences just as intelligible
28
29 as natural sentences, when heard in quiet).⁸ Presumably listeners used the remaining
30
31 phonetic information and their knowledge of Mandarin to guide their perception in a way
32
33 that allowed them to infer which words were being said. (Similar experiments have also
34
35 been performed in English.⁹) The larger point is that understanding the basic meaning of
36
37 sentences does not require ability to hear fine distinctions in pitch, due to the redundancy
38
39 of spoken language (this redundancy likely includes top-down semantic and syntactic
40
41 cues to word meaning). Of course, in natural language pitch variation does convey a
42
43 variety of information, such as prosodic focus, phrase boundaries, emotional tone,
44
45 attitude, and so forth. It appears, however, that deriving the basic semantic meaning of a
46
47 sentence does not seem to require high-precision pitch encoding by the brain.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In contrast, musical training does demand high-precision pitch encoding. This is because musical training involves constantly monitoring whether one is in tune and playing the right notes. In order to do this, a performer needs to hear subtle distinctions in pitch (in Western music, neighboring musical pitches, such as B vs. B-flat, differ by just ~6% in pitch). According to OPERA, this high-precision demand on pitch encoding sets the stage for musical training to sharpen pitch processing in the auditory system. The remaining three conditions of OPERA (Emotion, Repetition, and Attention) facilitate neural plasticity. According to OPERA, when all five conditions are met musical training drives subcortical pitch-encoding networks to function with higher precision than needed for ordinary speech processing. However, since speech and music share subcortical pitch-processing circuits, speech processing benefits. For example, higher-precision encoding of voice pitch may enhance speech perception in noise,^{10,11} since natural modulations in voice pitch contribute to enhanced intelligibility of speech in noise.^{8,9} Higher-precision encoding of voice pitch and may also contribute to the superior performance of musicians in recognizing affective prosody in spoken language.¹²

With this background in place, it is now possible to discuss certain assumptions and clarifications concerning the 'Precision' component of OPERA, focusing on pitch processing.

ASSUMPTION 1 : THE NEURAL ENCODING OF PITCH AND SPECTRAL SHAPE ARE DISTINCT

1
2
3 Speech perception relies on the neural encoding of several aspects of complex acoustic
4 signals. Important aspects include waveform periodicity (related to pitch), the
5
6 distribution of acoustic energy across the frequency spectrum, or spectral shape (related
7
8 to formant patterns and phoneme identity), and overall amplitude envelope (related to
9
10 syllable structure and prominence). In claiming that music demands greater precision in
11
12 pitch encoding than does speech, it is crucial to distinguish pitch from spectral shape.
13
14
15
16
17 Speech perception requires sensitivity to spectral shape and how it changes over time
18
19 (indeed, dynamic spectral shape information is both necessary and sufficient for speech
20
21 perception even in the absence of any pitch information).^{13,14} If the neural mechanisms
22
23 for encoding pitch and spectral shape were identical, then it would not make sense to
24
25 suggest that music demands greater auditory encoding precision than speech, because
26
27 speech places fairly high demands on the encoding of spectral shape (e.g., for detecting
28
29 formants and their movements). In other words, the idea that music demands higher-
30
31 precision pitch processing than speech depends on the assumption that the neural
32
33 encoding of pitch and of spectral shape rely on distinct brain mechanisms. For example,
34
35 the encoding of low-frequency pitches (such as those used in speech and music) may rely
36
37 heavily on temporal patterns of activity in auditory neurons, ultimately rooted in phase-
38
39 locking of action potentials to periodicities in the acoustic waveform.¹⁵ In contrast,
40
41 encoding of spectral shape may rely on spatial patterns of activity across auditory
42
43 tonotopic maps which occur in multiple subcortical and cortical regions. (As an aside, it
44
45 is interesting to note that humans can readily understand whispered speech, which
46
47 contains no acoustic periodicity and no clear pitch information, but which does contain
48
49 rapid spectral-shape changes due to acoustic filtering by vocal tract articulators. Hence
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the auditory system can do spectral shape encoding in the absence of any clear pitch
4
5 information).
6
7
8
9

10 **ASSUMPTION 2 : THE NORMAL AUDITORY SYSTEM TOLERATES ‘GOOD** 11 12 **ENOUGH’ PROCESSING** 13

14
15
16
17 If high-precision pitch encoding benefits speech perception (e.g., perhaps by enhancing
18 speech perception in noise and sensitivity to emotional prosody), then why doesn't our
19 day-to-day use of speech result in high-precision pitch encoding? OPERA assumes that
20 the precision of pitch encoding demanded by speech perception or any other kind of
21 auditory perception (e.g., perception of environmental sounds) reflects what is good
22 enough to get the job done. In ordinary speech perception, the primary job is
23 understanding the semantic content of utterances in the ecological contexts in which
24 spoken language most often takes place (i.e., in face-to-face conversations in relatively
25 quiet settings). If this is being done satisfactorily, then there is no “upward pressure” on
26 the auditory system to enhance the precision of pitch encoding. According to this view,
27 individuals can vary widely in their pitch encoding precision (perhaps due inborn
28 neuroanatomical differences) and still engage successfully in ordinary speech
29 communication, due to the many redundant cues to meaning in spoken language, as
30 discussed above. Crucially, however, when it comes to musical training the range of
31 encoding precision that is “good enough to get the job done” is narrower and skewed to
32 the high end. This is because musical training involves carefully regulating and
33 monitoring the pitches produced. Presumably, one cannot do this job with an auditory
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 system that does imprecise pitch encoding. This sets the stage for experience-dependent
4
5 plasticity to enhance pitch encoding by the brain.
6
7

8
9
10 **CLARIFICATION 1 : MUSICAL TRAINING AND SPEECH PERCEPTION –**
11
12 **APPLES AND ORANGES?**
13

14
15
16
17 OPERA is concerned with explaining why musical training benefits the neural encoding
18
19 of speech. OPERA compares the auditory encoding demands made by musical training
20
21 and by speech perception. These are very different activities. Most notably, musical
22
23 training involves sound production whereas speech perception does not. Hence does
24
25 OPERA compare apples and oranges, i.e., things that should not be compared? The
26
27 crucial point is that musical training also involves auditory training, because it involves
28
29 careful monitoring of what one produces. As mentioned earlier, musicians must
30
31 constantly listen to themselves to judge if they are in tune and are playing the right notes,
32
33 and these judgments require hearing subtle distinctions in pitch. Early in one's musical
34
35 training, feedback about whether one is in tune or playing the right notes may come from
36
37 an external source (such as a music teacher), but as one progresses self-monitoring
38
39 becomes essential, and production and perception become tightly intertwined. Thus in
40
41 discussing the auditory encoding demands made by musical training vs. by speech
42
43 perception, one is really talking about the demands made by music perception (in the
44
45 context of performance training) vs. speech perception.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CLARIFICATION 2 : MUSICAL PERFORMANCE TRAINING VS. EAR TRAINING

Musical training does not always involve sound production. ‘Ear training’ is also part of music education, e.g., learning to distinguishing different rhythms, pitch intervals, cadence types, and so forth. In principle, if such ear training requires making subtle pitch distinctions and also involves positive emotion, extensive repetition, and focused attention, then purely aural training in music could enhance the neural encoding of pitch in speech according to the principles of OPERA. However, in practice it may be much easier to meet the conditions of OPERA in performance-based musical training than in purely perception-based training. For example, one major difference between performance-based training and ear-training is that the former is often social, involving group music-making. This can heighten emotion and attention due to the psychological impact of a engaging in a coordinated, pleasing group activity. Furthermore, sensorimotor training may drive neural plasticity more strongly than purely sensory training.¹⁶

LOOKING AHEAD : BEYOND PITCH

Aside from pitch, are there other auditory attributes for which musical training might demand higher precision encoding than speech perception? Of course, what constitutes “musical training” depends on the instrument and the type of music being studied. For example, there are many percussion instruments and musical styles that place a high

1
2
3 premium on rhythm and timing. How do rhythmic music and speech compare in terms of
4 the demands placed on accurate encoding of timing information? Small temporal
5 differences can be quite important in speech. For example, small differences in voice-
6 onset-time help distinguishing stop consonants such as /b/ vs. /p/, and small differences in
7 the rise time of amplitude envelopes help distinguish stop consonants from glides, such as
8 /b/ vs. /w/. However, just as with pitch, the critical question is “how important is hearing
9 fine timing distinctions for speech comprehension vs. for musical training?” It may be
10 that due to the redundancy of spoken language, if a person cannot hear fine temporal
11 distinctions he or she will still comprehend the basic meaning of sentences. In contrast,
12 musical training on rhythmic instruments may require hearing fine timing distinctions,
13 e.g., as part of determining whether a rhythm is being played correctly or whether one is
14 in synchrony with other members of a group.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31
32 More broadly, empirical work is needed to determine which auditory attributes
33 shared by speech and music are subject to higher-precision encoding demands in the
34 musical domain. Such information can help guide hypothesis-driven explorations of
35 music as a tool for enhancing or rehabilitating certain speech perception abilities.
36
37
38
39
40
41
42

43 **ACKNOWLEDGMENTS**

44
45
46
47
48 I thank D. Robert Ladd, LouAnn Gerken, Devin McAuley, and Dan Levitin for insightful
49 comments on the original OPERA paper, which helped inspire this update, and Mathias
50 Oechslin, Nina Kraus, Trent Nicol, and Alexandra Parbery-Clark for comments on this
51 paper. Supported by Neurosciences Research Foundation as part of its program on music
52
53
54
55
56
57
58
59
60

1
2
3 and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham Senior
4
5
6 Fellow.

7
8
9
10 **REFERENCES**

- 11
12
13
14
15 1. Kraus, N. & B. Chandrasekaran. 2010. Music training for the development of auditory
16
17 skills. *Nat. Rev. Neurosci.* 11:599-605.
18
19
20
21
22
23 2. Patel, A.D., & J.R. Iversen. 2007. The linguistic benefits of musical abilities. *Trends*
24
25 *in Cognitive Sciences* 11: 369-372.
26
27
28
29
30 3. Song, J. H., E. Skoe, P. C. Wong, & N. Kraus. 2008. Plasticity in the adult human
31
32 auditory brainstem following short-term linguistic training. *Journal of Cognitive Neurosci.*
33
34 10: 1892–1902.
35
36
37
38
39 4. Moreno, S., C. Marques, A. Santos, et al. 2009. Musical training influences linguistic
40
41 abilities in 8-Year-Old children: More evidence for brain plasticity. *Cerebral Cortex* 19:
42
43 712-723.
44
45
46
47
48 5. Patel, A.D. 2011. Why would musical training benefit the neural encoding of speech?
49
50 The OPERA hypothesis. *Frontiers in Psychology* 2:142. (doi: 10.3389/
51
52 fpsyg.2011.00142)
53
54
55
56
57
58
59
60

- 1
2
3 6. Patel, A.D. 2008. Music, Language, and the Brain. Oxford Univ. Press. New York.
4
5
- 6
7
8 7. McDermott, J.H., & A.J. Oxenham. 2008. Music perception, pitch, and the auditory
9
10 system. *Current Opinion in Neurobiology* 18: 452–463.
11
12
- 13
14
15 8. Patel, A.D., Y. Xu, & B. Wang. 2010. The role of F0 variation in the intelligibility of
16
17 Mandarin sentences. *Proceedings of Speech Prosody 2010*, May 11-14, 2010, Chicago,
18
19 IL, USA.
20
21
22
- 23
24
25 9. Binns, C., & Culling, J.F. 2008. The role of fundamental frequency contours in the
26
27 perception of speech against interfering speech. *J. Acoust. Soc. Am.* 122: 1765-1776.
28
29
- 30
31
32 10. Parbery-Clark, A., Skoe, E. & Kraus, N. 2009. Musical experience limits the
33
34 degradative effects of background noise on the neural processing of sound. *J. Neurosci.*
35
36 29: 14100–14107.
37
38
- 39
40
41 11. Parbery-Clark, A., D.L. Strait, S. Anderson, E. Hittner, & N. Kraus. 2011.
42
43 Musical training and the aging auditory system: implications for cognitive abilities and
44
45 hearing speech in noise. *PLoS ONE* 6: e18082. doi: 10.1371/ journal.pone.0018082
46
47
48
- 49
50
51 12. Lima, C.F. & Castro, S.L. 2011. Speaking to the trained ear: musical expertise
52
53 enhances the recognition of emotions in speech prosody. *Emotion* 11: 1021-1031.
54
55
56
57
58
59
60

1
2
3 13. Drullman, R., J.M. Festen, & R. Plomp. 1994. Effect of temporal envelope smearing
4 on speech reception. *J. Acoust. Soc. Am.* 95:1053–1064.
5
6
7

8
9
10 14. Shannon, R.V., F.G. Zeng, V. Kamath, J. Wygonski, & M. Ekelid. 1995. Speech
11 recognition with primarily temporal cues. *Science* 270:303–304.
12
13
14

15
16
17 15. Cariani, P. A. & B. Delgutte, B. 1996. Neural correlates of the pitch of complex tones.
18 I. Pitch and pitch salience. *J. Neurophysiol.* 76: 1698–1716.
19
20
21

22
23
24 16. Lappe, C., S.C. Herholz, L.J. Trainor, & C. Pantev. 2008. Cortical plasticity induced
25 by short-term unimodal and multimodal musical training. *J. Neurosci.* 8, 9632–9639.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60