UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Source Coding Project

LBG-split for coding and decoding of CD-quality audio signals

Student:
Umberto Michieli

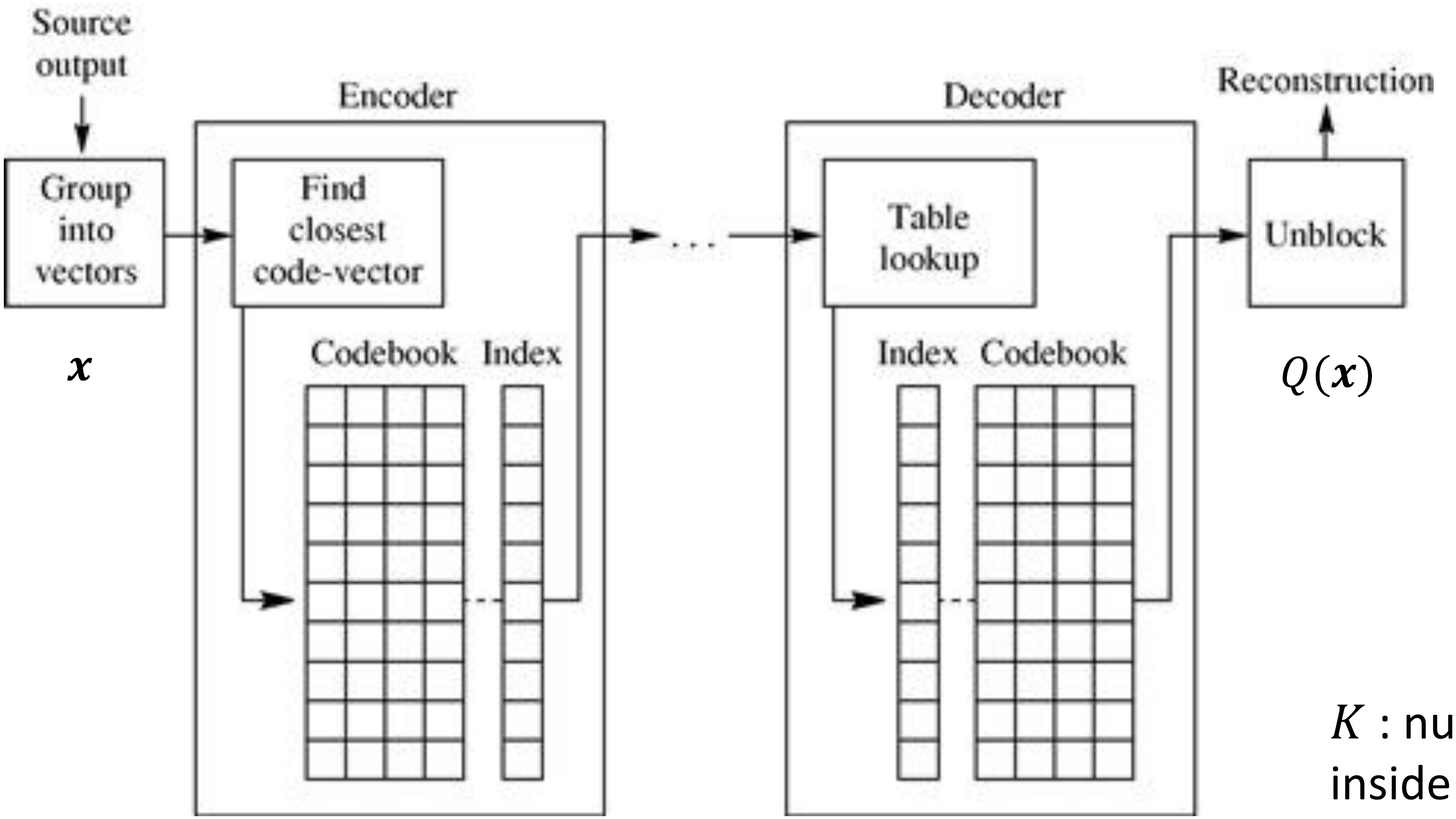
19 luglio 2017

Introduction to Coding Tecniques

- **Lossless Coding**: invertible (no loss of information)
  exploit variable-length coding
  compression ratio small

- **Lossy Coding**: not invertible (loss of information)
  main idea is quantization
  -minimum rate for a given distortion
  -minimum distortion given the rate

Vector Quantization (VQ)



$\boldsymbol{x} \in \mathbb{R}^L$

Partition of $\mathbb{R}^L$:

$I_i \subseteq \mathbb{R}^L, i = 1, \dots, K$

$I_i \cap I_j = \emptyset, \forall\, i \neq j$

$$\bigcup_{i=1}^{K} I_i = \mathbb{R}^L$$

K : number of codevectors inside the codebook

$R = \frac{1}{L}\lceil log_2 K \rceil$ bits/component

$$SNR = \frac{\sigma^2_{reconstructed\_signal}}{\sigma^2_{quantization\_noise}}$$

LBG: Algorithm with pdf unknown

1) Inizialization: given T, codebook $\{y_1^{(0)}, \dots, y_K^{(0)}\}$, $n = 1$, $D^{(0)} = \infty$, $\varepsilon > 0$

   Where T is the training set, $D^{(0)}$ is the initial distortion and ε a termination threshold

2) Optimal partitioning: (*Nearest Neighbour Condition*):

$$I_i^{(n)} = \{\mathbf{x} \in \mathcal{T} \text{ such that } ||\mathbf{x} - \mathbf{y}_i^{(n-1)}||_2^2 \leq ||\mathbf{x} - \mathbf{y}_j^{(n-1)}||_2^2,\ i \neq j\},\ i = 1, \dots, K$$

3) New codebook (Centroid Condition):

$$\mathbf{y}_i^{(n)} = \frac{1}{|I_i^{(n)}|} \sum_{\mathbf{x} \in I_i^{(n)}} \mathbf{x}$$

4) Total distortion:

$$D^{(n)} = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} ||\mathbf{x} - Q(\mathbf{x})||_2^2,$$

5) If $\dfrac{D^{(n-1)} - D^{(n)}}{D^{(n)}} < \varepsilon$ then stop,

   else $n \leftarrow n + 1$ and go to step 2

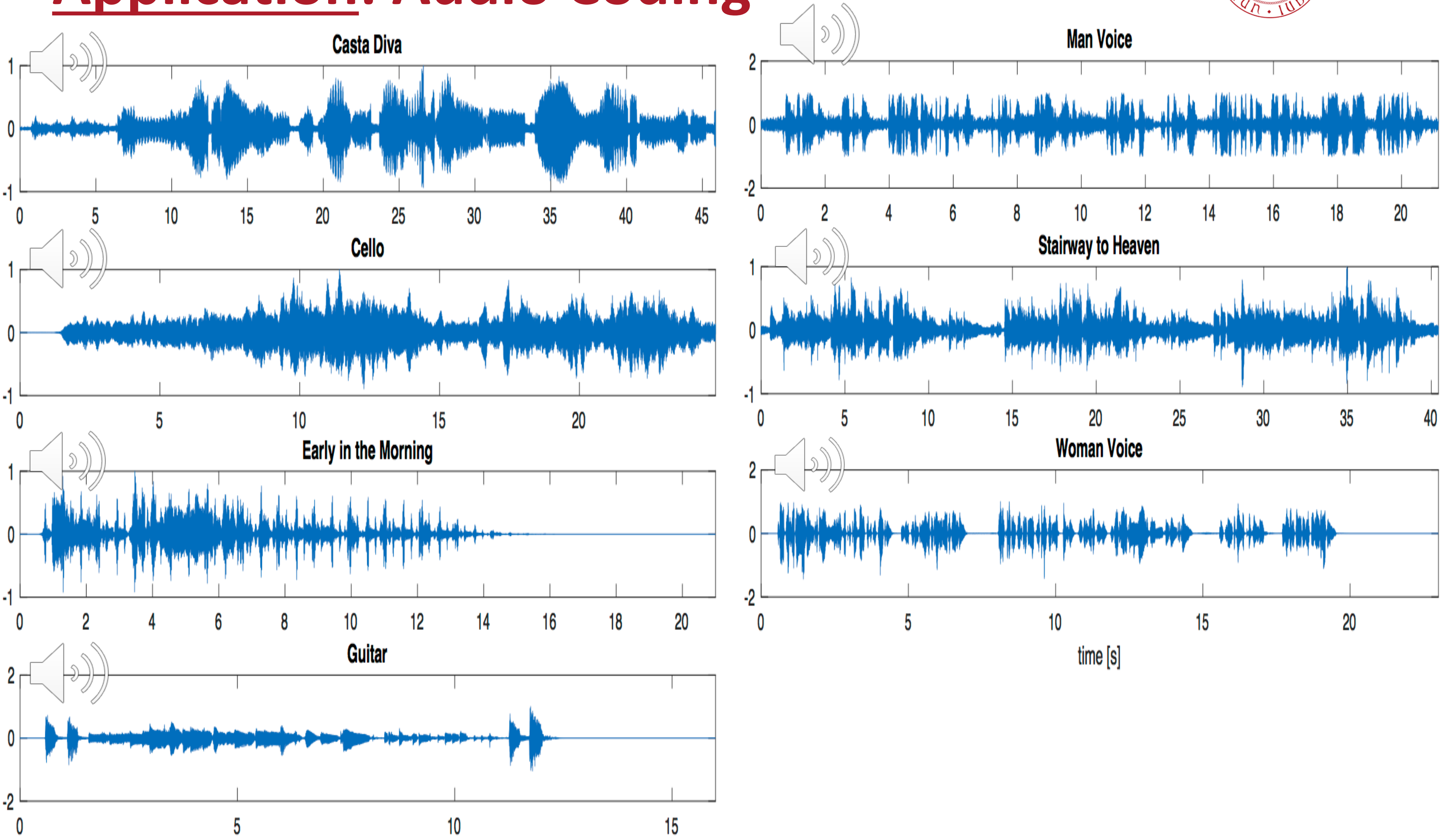
LBG: Discussion

- No guarantee of a global minimum of the distortion (local minima are possible)
  → it depends on the initial codebook

- Codebook initialization: **splitting technique**
  - VQ with a single output point → average value of the entire training set
  - 2-level VQ adding and removing a perturbation $= 0.1$ to the codevectors
  - Iteration until the number of codevectors is *K*

- Termination threshold set to: $\varepsilon = 0.1$

Application: Audio Coding



Results: L=2, mono, training=*Casta Diva*

Audio name	*L*	*K*	*R bit/sample*	SNR$_{\text{dB}}$	*K*	*R bit/sample*	SNR$_{\text{dB}}$
Casta Diva	**2**	**2**	**0.5**	**2.8725**	**4**	**1**	**7.4876**
Cello	2	2	0.5	2.6925	4	1	7.3330
Early in the Morning	2	2	0.5	0.8704	4	1	5.6790
Guitar	2	2	0.5	-1.8154	4	1	4.3871
Man Voice	2	2	0.5	1.4508	4	1	3.2558
Stairway to Heaven	2	2	0.5	-1.9086	4	1	4.5466
Woman Voice	2	2	0.5	1.3290	4	1	6.1607
Casta Diva	**2**	**8**	**1.5**	**12.9501**	**16**	**2**	**17.8592**
Cello	2	8	1.5	12.5083	16	2	17.6690
Early in the Morning	2	8	1.5	10.0789	16	2	13.4206
Guitar	2	8	1.5	10.2144	16	2	15.5869
Man Voice	2	8	1.5	4.9777	16	2	6.1117
Stairway to Heaven	2	8	1.5	10.0850	16	2	15.3314
Woman Voice	2	8	1.5	10.3523	16	2	13.1654
Casta Diva	**2**	**32**	**2.5**	**21.9642**	**64**	**3**	**24.6008**
Cello	2	32	2.5	22.0293	64	3	24.9176
Early in the Morning	2	32	2.5	15.7615	64	3	17.2942
Guitar	2	32	2.5	19.3825	64	3	21.2062
Man Voice	2	32	2.5	6.9639	64	3	7.6450
Stairway to Heaven	2	32	2.5	19.2954	64	3	21.2993
Woman Voice	2	32	2.5	14.6991	64	3	15.4456
Casta Diva	**2**	**128**	**3.5**	**25.9558**	**256**	**4**	**29.6845**
Cello	2	128	3.5	26.5651	256	4	29.8972
Early in the Morning	2	128	3.5	18.6354	256	4	21.1628
Guitar	2	128	3.5	21.8688	256	4	23.9498
Man Voice	2	128	3.5	8.4725	256	4	9.4948
Stairway to Heaven	2	128	3.5	22.0431	256	4	25.8277
Woman Voice	2	128	3.5	15.9338	256	4	17.2472

✓ As *K* increases, *SNR* increases, but also *R* increases (can be heard)

✓ *Casta Diva*'s SNR is generally greater than the others

✓ Anomaly: *Cello*

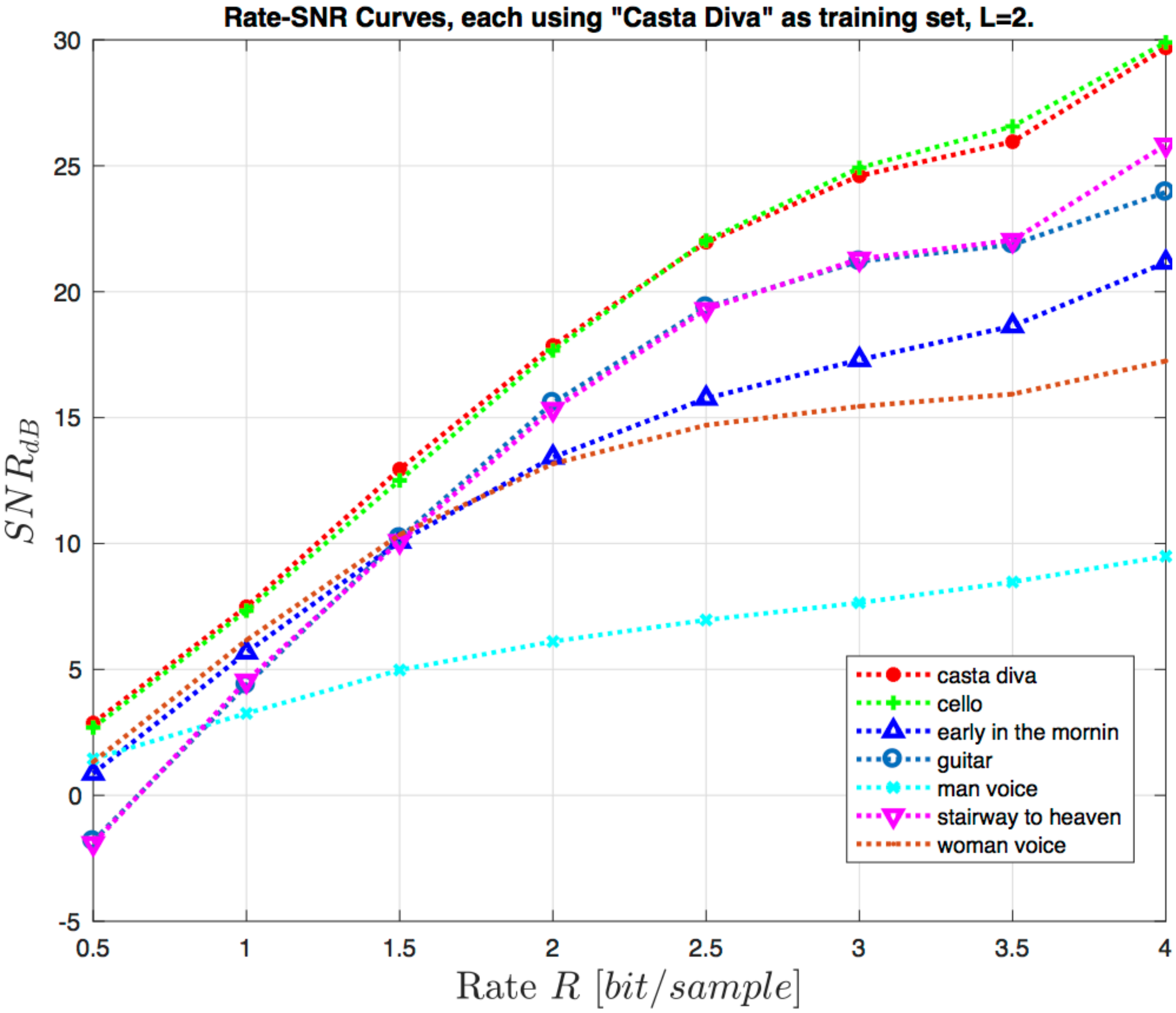
Why?

$-var(Cello) \approx 2\ var(Casta\ Diva)$

-samples of *Cello* closer to the codevectors in mean sense

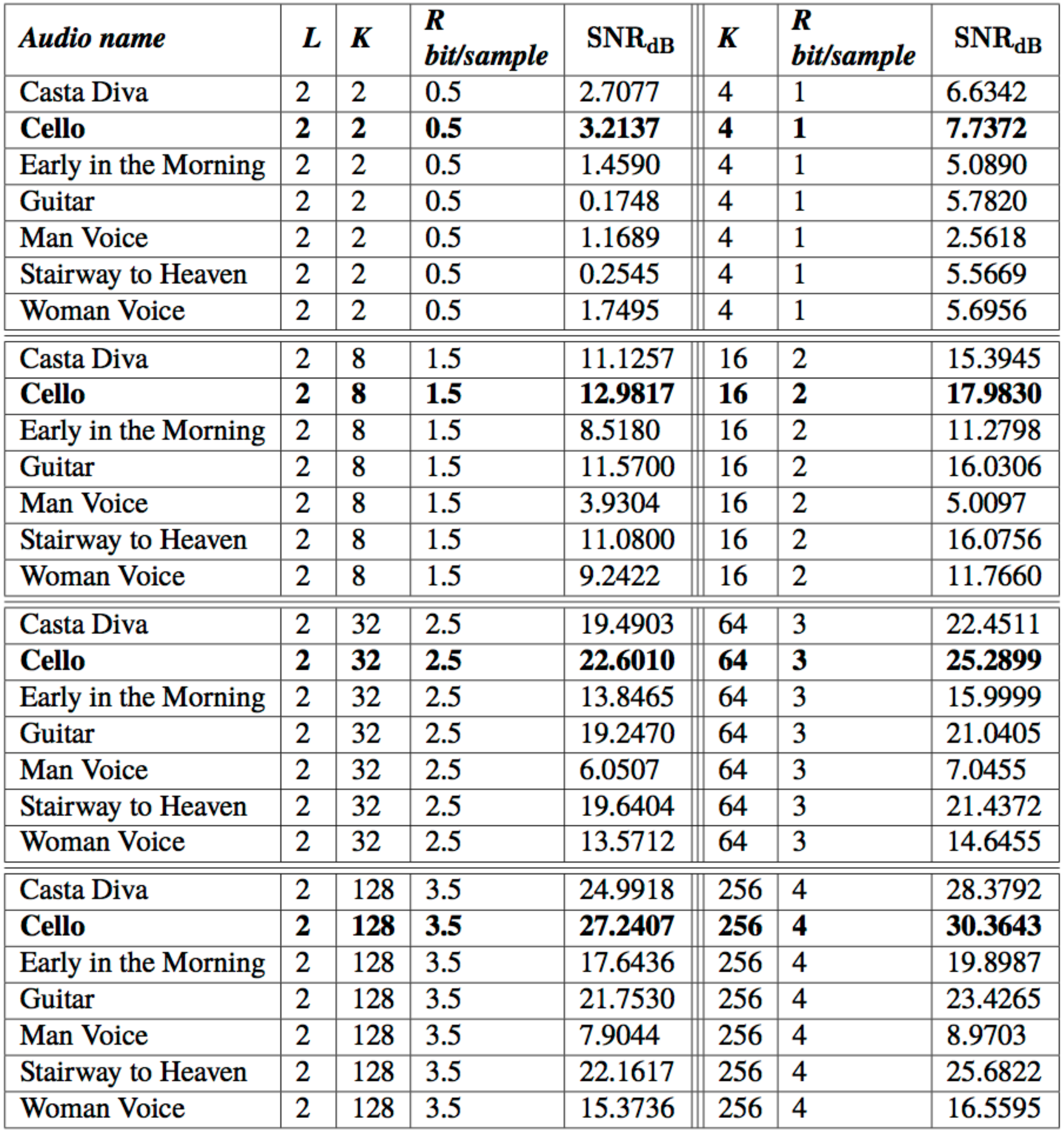
Results: L=2, mono, training=*Casta Diva*

Università degli Studi di Padova



- Can match the points with the table
- *Casta Diva*'s SNR is generally greater than the others
- Anomaly: *Cello*
- *Man voice* cannot be well-represented by a woman singing opera

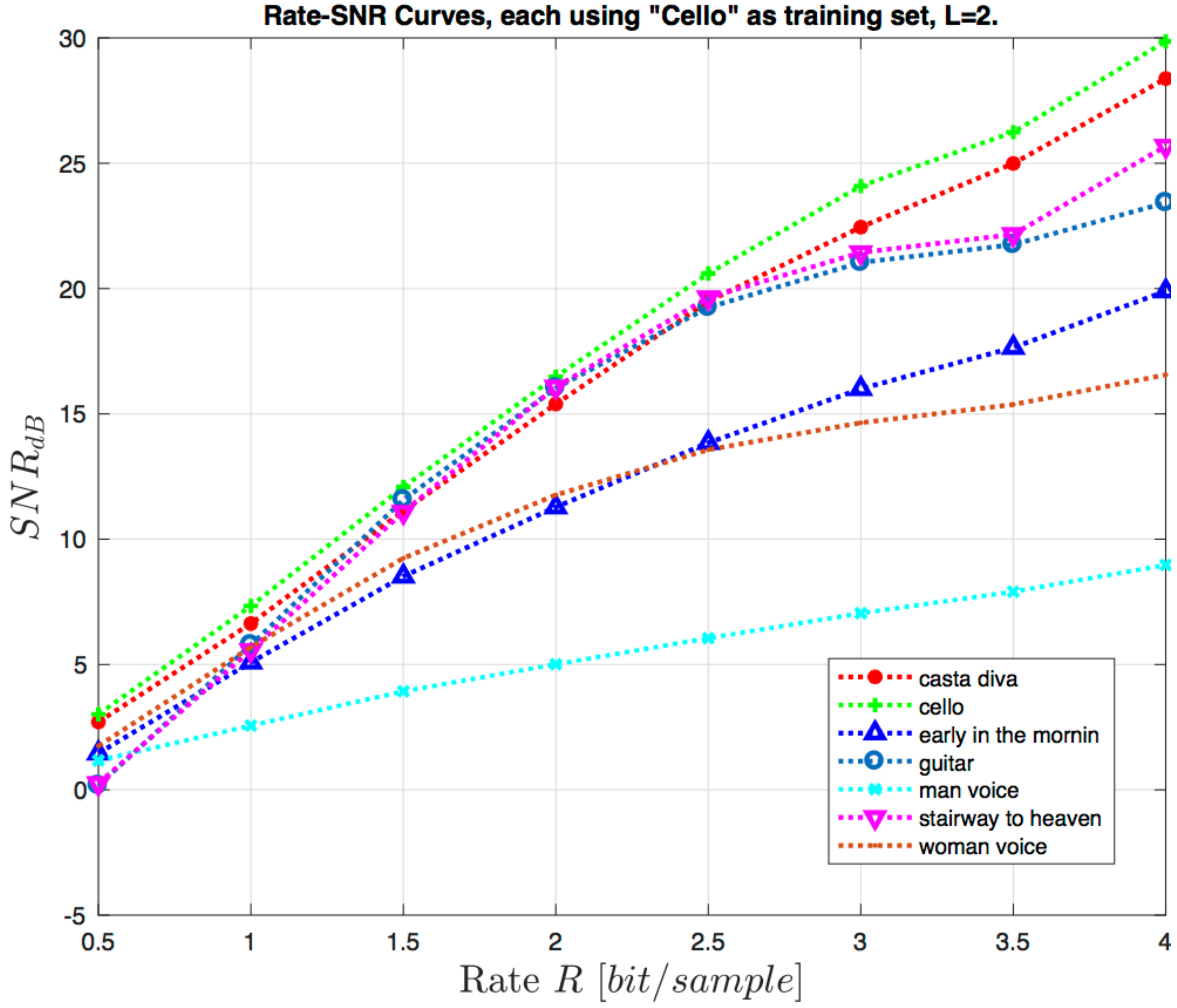
Results: L=2, mono, training=*Cello*



Audio name	*L*	*K*	*R bit/sample*	SNR$_{dB}$	*K*	*R bit/sample*	SNR$_{dB}$
Casta Diva	2	2	0.5	2.7077	4	1	6.6342
Cello	**2**	**2**	**0.5**	**3.2137**	**4**	**1**	**7.7372**
Early in the Morning	2	2	0.5	1.4590	4	1	5.0890
Guitar	2	2	0.5	0.1748	4	1	5.7820
Man Voice	2	2	0.5	1.1689	4	1	2.5618
Stairway to Heaven	2	2	0.5	0.2545	4	1	5.5669
Woman Voice	2	2	0.5	1.7495	4	1	5.6956
Casta Diva	2	8	1.5	11.1257	16	2	15.3945
Cello	**2**	**8**	**1.5**	**12.9817**	**16**	**2**	**17.9830**
Early in the Morning	2	8	1.5	8.5180	16	2	11.2798
Guitar	2	8	1.5	11.5700	16	2	16.0306
Man Voice	2	8	1.5	3.9304	16	2	5.0097
Stairway to Heaven	2	8	1.5	11.0800	16	2	16.0756
Woman Voice	2	8	1.5	9.2422	16	2	11.7660
Casta Diva	2	32	2.5	19.4903	64	3	22.4511
Cello	**2**	**32**	**2.5**	**22.6010**	**64**	**3**	**25.2899**
Early in the Morning	2	32	2.5	13.8465	64	3	15.9999
Guitar	2	32	2.5	19.2470	64	3	21.0405
Man Voice	2	32	2.5	6.0507	64	3	7.0455
Stairway to Heaven	2	32	2.5	19.6404	64	3	21.4372
Woman Voice	2	32	2.5	13.5712	64	3	14.6455
Casta Diva	2	128	3.5	24.9918	256	4	28.3792
Cello	**2**	**128**	**3.5**	**27.2407**	**256**	**4**	**30.3643**
Early in the Morning	2	128	3.5	17.6436	256	4	19.8987
Guitar	2	128	3.5	21.7530	256	4	23.4265
Man Voice	2	128	3.5	7.9044	256	4	8.9703
Stairway to Heaven	2	128	3.5	22.1617	256	4	25.6822
Woman Voice	2	128	3.5	15.3736	256	4	16.5595

- ✓ *SNR* for *Cello* slightly higher than *Casta Diva*
- ✓ *SNR* for *Cello* slightly higher than before

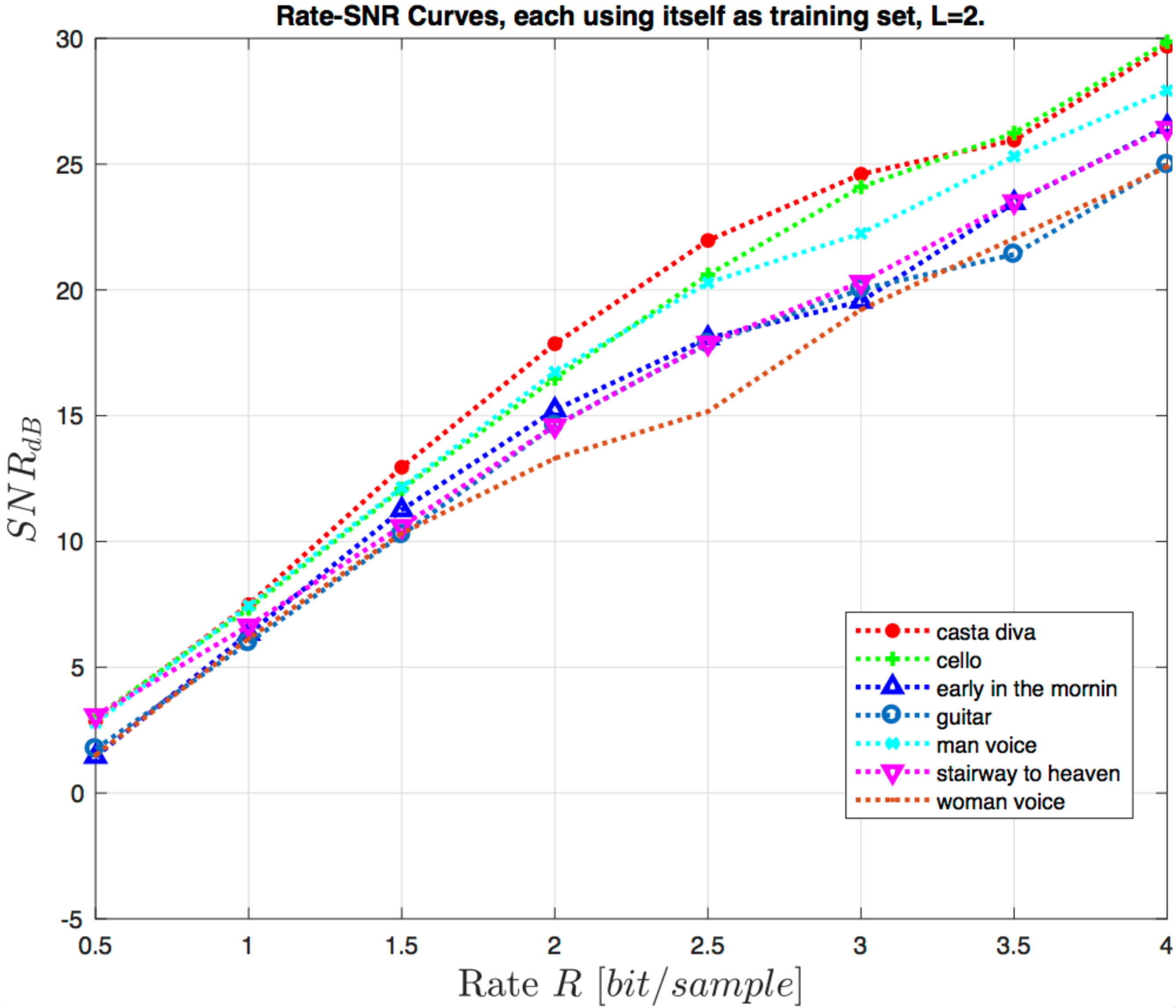
Results: L=2, mono, training=*Cello*



- ✓ *SNR* for *Cello* slightly higher than *Casta Diva*
- ✓ *SNR* for *Cello* slightly higher than before

Results: L=2, mono, training=*itself*

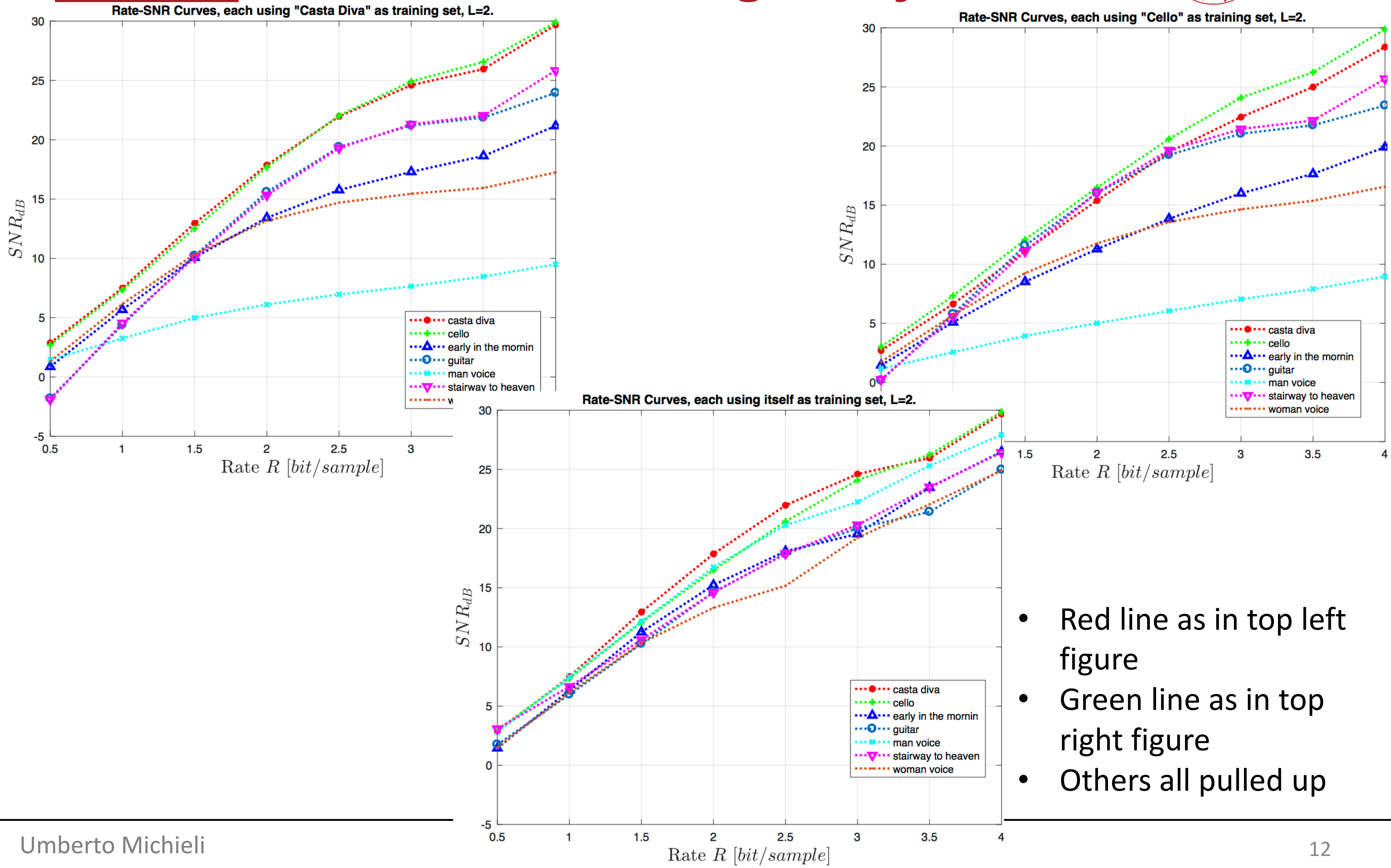
UNIVERSITÀ DEGLI STUDI DI PADOVA



- Need to transmit the codebook

✓ Every *SNR*-curve has been pulled up

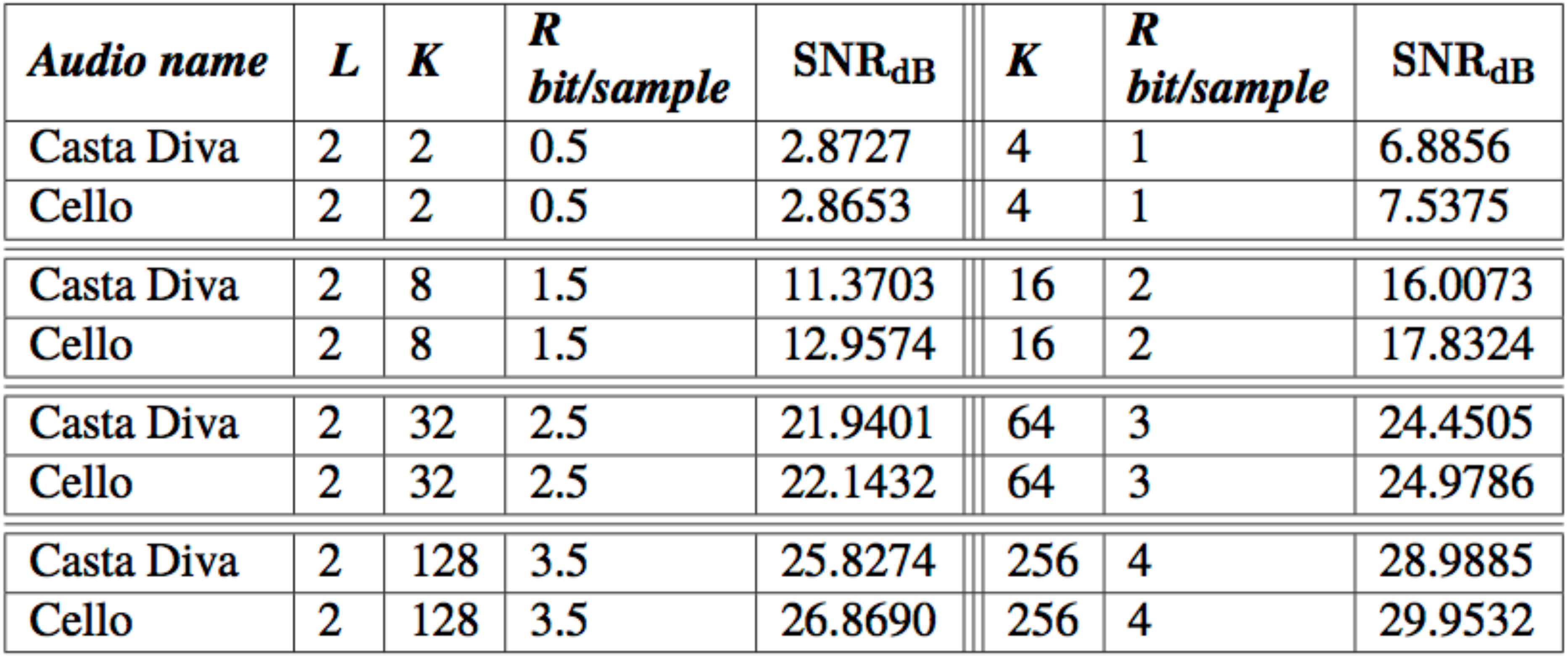
Results: L=2, mono, training=*itself*



- Red line as in top left figure
- Green line as in top right figure
- Others all pulled up

Results: L=2, mono, training=*Mixed1*

Mixed1 is composed by pieces of audio of the signals to code

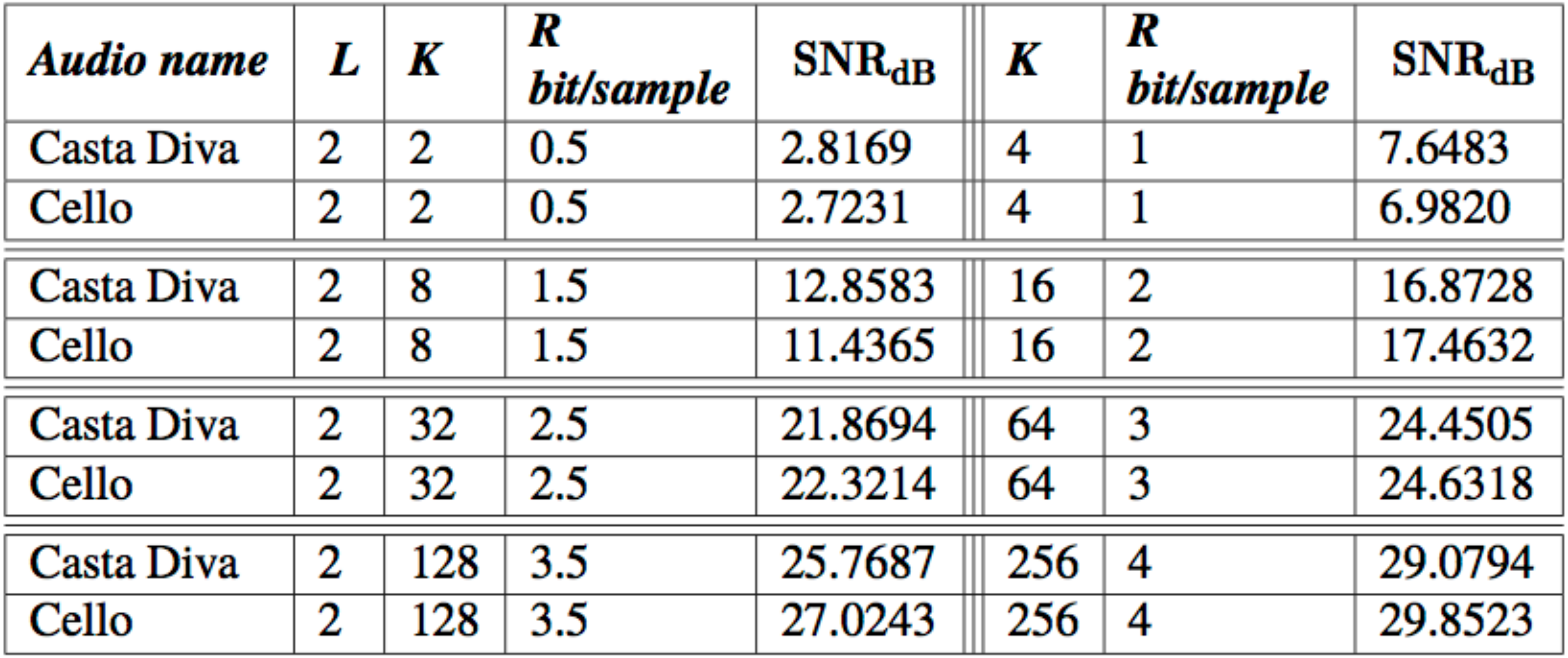


Audio name	*L*	*K*	*R* *bit/sample*	SNR$_{dB}$	*K*	*R* *bit/sample*	SNR$_{dB}$
Casta Diva	2	2	0.5	2.8727	4	1	6.8856
Cello	2	2	0.5	2.8653	4	1	7.5375
Casta Diva	2	8	1.5	11.3703	16	2	16.0073
Cello	2	8	1.5	12.9574	16	2	17.8324
Casta Diva	2	32	2.5	21.9401	64	3	24.4505
Cello	2	32	2.5	22.1432	64	3	24.9786
Casta Diva	2	128	3.5	25.8274	256	4	28.9885
Cello	2	128	3.5	26.8690	256	4	29.9532

Values very similar as before → does it depend on *Mixed1*?

Results: L=2, mono, training=*Mixed2*

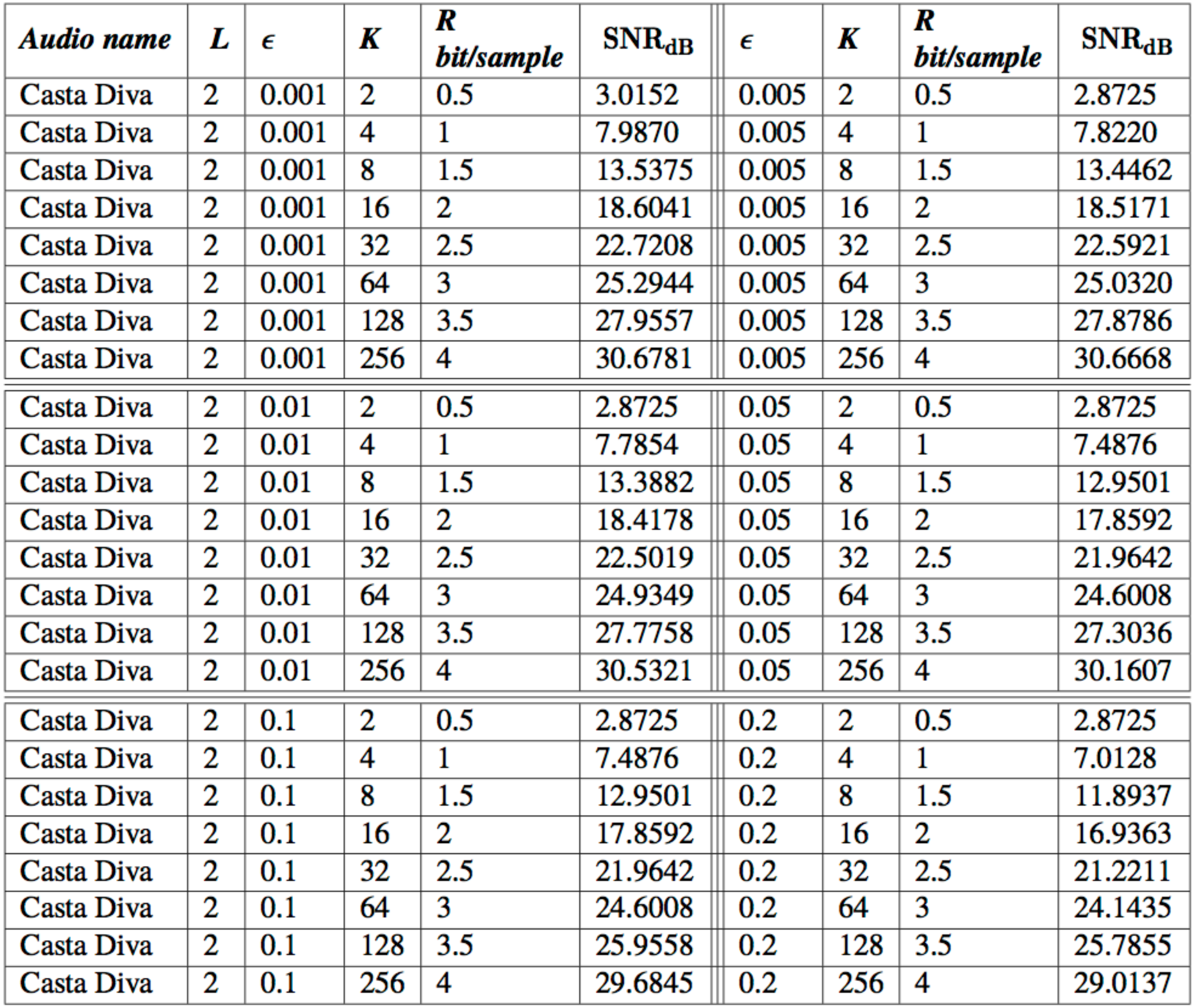
Mixed2 is composed using various pieces of audios



Audio name	***L***	***K***	***R*** ***bit/sample***	SNR$_{dB}$	***K***	***R*** ***bit/sample***	SNR$_{dB}$
Casta Diva	2	2	0.5	2.8169	4	1	7.6483
Cello	2	2	0.5	2.7231	4	1	6.9820
Casta Diva	2	8	1.5	12.8583	16	2	16.8728
Cello	2	8	1.5	11.4365	16	2	17.4632
Casta Diva	2	32	2.5	21.8694	64	3	24.4505
Cello	2	32	2.5	22.3214	64	3	24.6318
Casta Diva	2	128	3.5	25.7687	256	4	29.0794
Cello	2	128	3.5	27.0243	256	4	29.8523

Values very similar as before → the distortion introduced by the LBG mostly depends on the input audio, not on the training set (unless the signal itself is used)

Results: L=2, mono, training=*Casta Diva*



Audio name	*L*	ϵ	*K*	*R* *bit/sample*	$\mathrm{SNR}_{\mathrm{dB}}$	ϵ	*K*	*R* *bit/sample*	$\mathrm{SNR}_{\mathrm{dB}}$
Casta Diva	2	0.001	2	0.5	3.0152	0.005	2	0.5	2.8725
Casta Diva	2	0.001	4	1	7.9870	0.005	4	1	7.8220
Casta Diva	2	0.001	8	1.5	13.5375	0.005	8	1.5	13.4462
Casta Diva	2	0.001	16	2	18.6041	0.005	16	2	18.5171
Casta Diva	2	0.001	32	2.5	22.7208	0.005	32	2.5	22.5921
Casta Diva	2	0.001	64	3	25.2944	0.005	64	3	25.0320
Casta Diva	2	0.001	128	3.5	27.9557	0.005	128	3.5	27.8786
Casta Diva	2	0.001	256	4	30.6781	0.005	256	4	30.6668
Casta Diva	2	0.01	2	0.5	2.8725	0.05	2	0.5	2.8725
Casta Diva	2	0.01	4	1	7.7854	0.05	4	1	7.4876
Casta Diva	2	0.01	8	1.5	13.3882	0.05	8	1.5	12.9501
Casta Diva	2	0.01	16	2	18.4178	0.05	16	2	17.8592
Casta Diva	2	0.01	32	2.5	22.5019	0.05	32	2.5	21.9642
Casta Diva	2	0.01	64	3	24.9349	0.05	64	3	24.6008
Casta Diva	2	0.01	128	3.5	27.7758	0.05	128	3.5	27.3036
Casta Diva	2	0.01	256	4	30.5321	0.05	256	4	30.1607
Casta Diva	2	0.1	2	0.5	2.8725	0.2	2	0.5	2.8725
Casta Diva	2	0.1	4	1	7.4876	0.2	4	1	7.0128
Casta Diva	2	0.1	8	1.5	12.9501	0.2	8	1.5	11.8937
Casta Diva	2	0.1	16	2	17.8592	0.2	16	2	16.9363
Casta Diva	2	0.1	32	2.5	21.9642	0.2	32	2.5	21.2211
Casta Diva	2	0.1	64	3	24.6008	0.2	64	3	24.1435
Casta Diva	2	0.1	128	3.5	25.9558	0.2	128	3.5	25.7855
Casta Diva	2	0.1	256	4	29.6845	0.2	256	4	29.0137

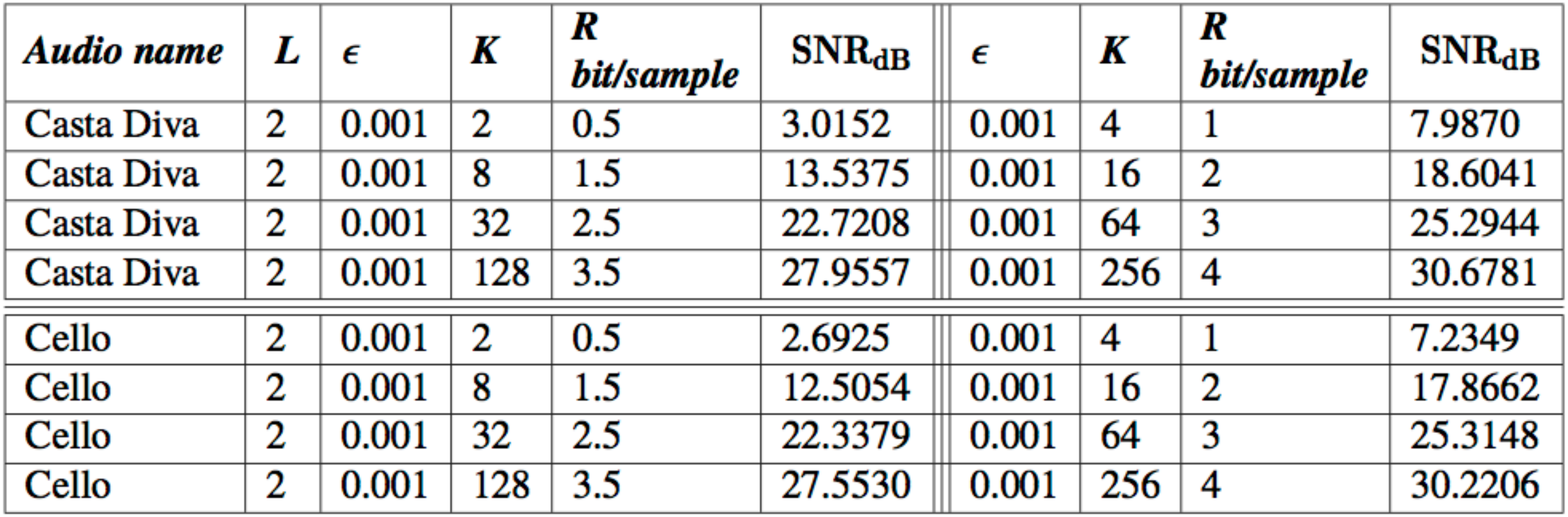
Up to now: $\varepsilon = 0.1$

✓ As ε decreases the *SNR* increases (closer to the codevectors)

- As ε decreases the complexity increases (more iterations)

Results: L=2, mono, training=*Casta Diva*

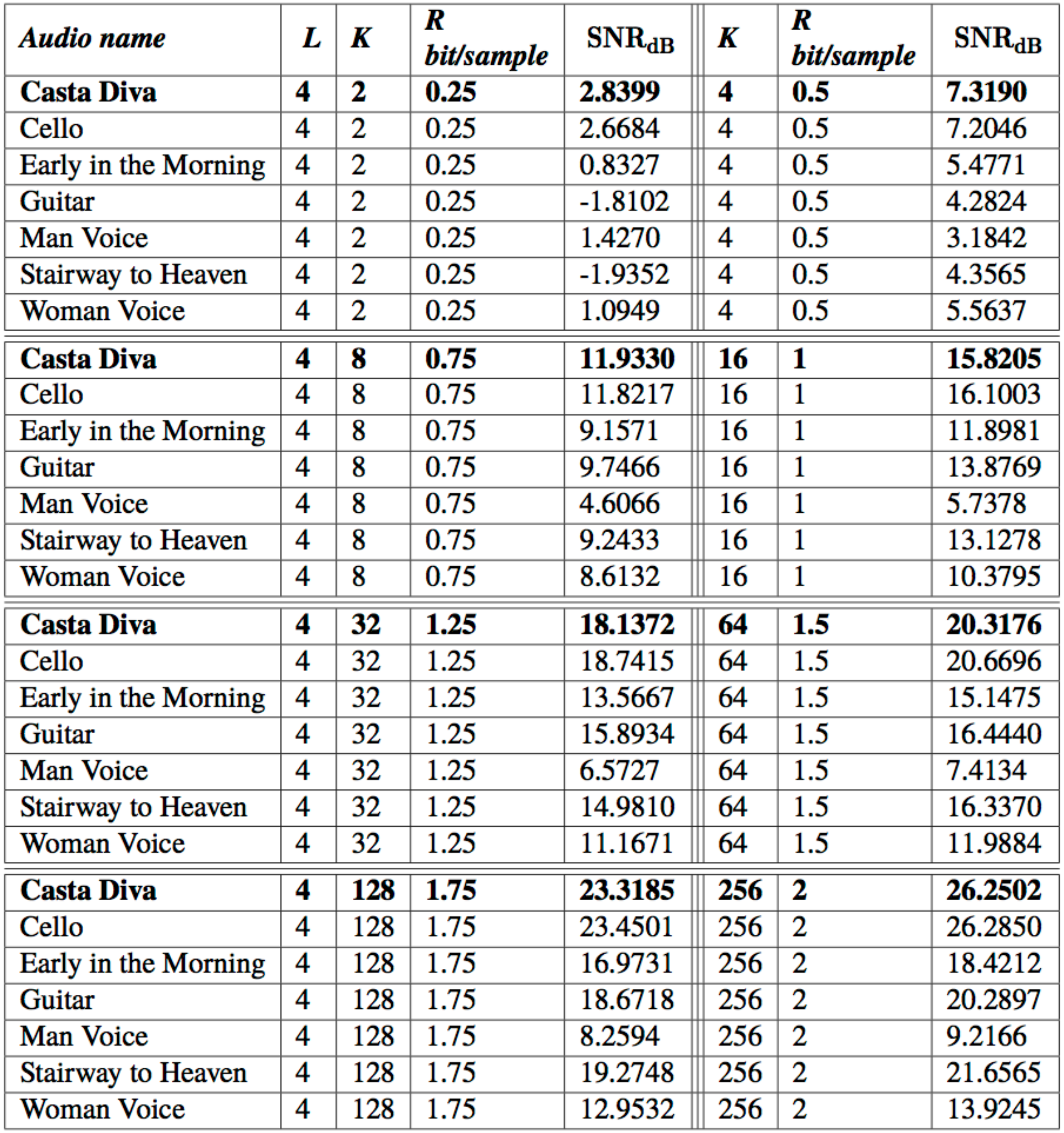
What about the *anomaly* of *Cello*?



Audio name	*L*	ϵ	*K*	*R* *bit/sample*	$\mathrm{SNR}_{\mathrm{dB}}$	ϵ	*K*	*R* *bit/sample*	$\mathrm{SNR}_{\mathrm{dB}}$
Casta Diva	2	0.001	2	0.5	3.0152	0.001	4	1	7.9870
Casta Diva	2	0.001	8	1.5	13.5375	0.001	16	2	18.6041
Casta Diva	2	0.001	32	2.5	22.7208	0.001	64	3	25.2944
Casta Diva	2	0.001	128	3.5	27.9557	0.001	256	4	30.6781
Cello	2	0.001	2	0.5	2.6925	0.001	4	1	7.2349
Cello	2	0.001	8	1.5	12.5054	0.001	16	2	17.8662
Cello	2	0.001	32	2.5	22.3379	0.001	64	3	25.3148
Cello	2	0.001	128	3.5	27.5530	0.001	256	4	30.2206

Now *Cello* has an higher *SNR* than *Casta Diva*, as it should be

Results: L=4, mono, training=*Casta Diva*



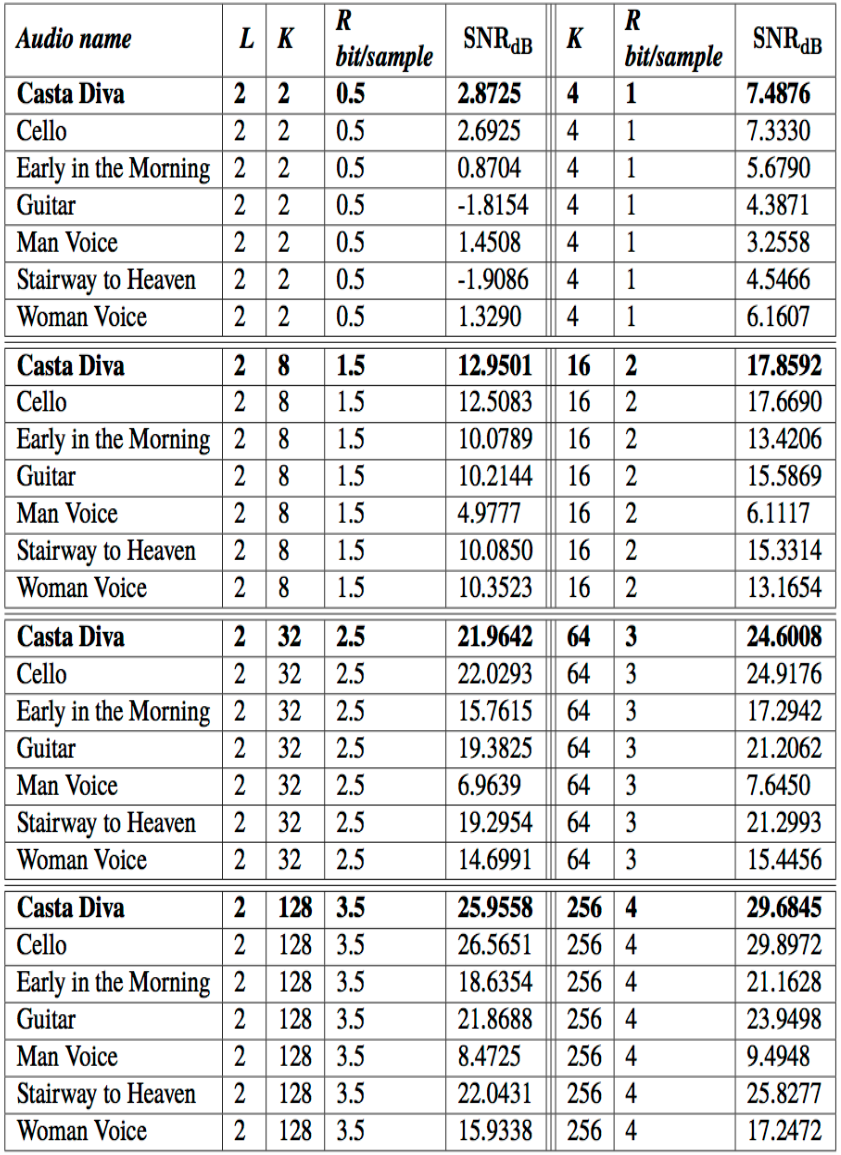
Audio name	*L*	*K*	*R* *bit/sample*	SNR_{dB}	*K*	*R* *bit/sample*	SNR_{dB}
Casta Diva	**4**	**2**	**0.25**	**2.8399**	**4**	**0.5**	**7.3190**
Cello	4	2	0.25	2.6684	4	0.5	7.2046
Early in the Morning	4	2	0.25	0.8327	4	0.5	5.4771
Guitar	4	2	0.25	-1.8102	4	0.5	4.2824
Man Voice	4	2	0.25	1.4270	4	0.5	3.1842
Stairway to Heaven	4	2	0.25	-1.9352	4	0.5	4.3565
Woman Voice	4	2	0.25	1.0949	4	0.5	5.5637
Casta Diva	**4**	**8**	**0.75**	**11.9330**	**16**	**1**	**15.8205**
Cello	4	8	0.75	11.8217	16	1	16.1003
Early in the Morning	4	8	0.75	9.1571	16	1	11.8981
Guitar	4	8	0.75	9.7466	16	1	13.8769
Man Voice	4	8	0.75	4.6066	16	1	5.7378
Stairway to Heaven	4	8	0.75	9.2433	16	1	13.1278
Woman Voice	4	8	0.75	8.6132	16	1	10.3795
Casta Diva	**4**	**32**	**1.25**	**18.1372**	**64**	**1.5**	**20.3176**
Cello	4	32	1.25	18.7415	64	1.5	20.6696
Early in the Morning	4	32	1.25	13.5667	64	1.5	15.1475
Guitar	4	32	1.25	15.8934	64	1.5	16.4440
Man Voice	4	32	1.25	6.5727	64	1.5	7.4134
Stairway to Heaven	4	32	1.25	14.9810	64	1.5	16.3370
Woman Voice	4	32	1.25	11.1671	64	1.5	11.9884
Casta Diva	**4**	**128**	**1.75**	**23.3185**	**256**	**2**	**26.2502**
Cello	4	128	1.75	23.4501	256	2	26.2850
Early in the Morning	4	128	1.75	16.9731	256	2	18.4212
Guitar	4	128	1.75	18.6718	256	2	20.2897
Man Voice	4	128	1.75	8.2594	256	2	9.2166
Stairway to Heaven	4	128	1.75	19.2748	256	2	21.6565
Woman Voice	4	128	1.75	12.9532	256	2	13.9245

At the same *K* as in *L=2* we have:

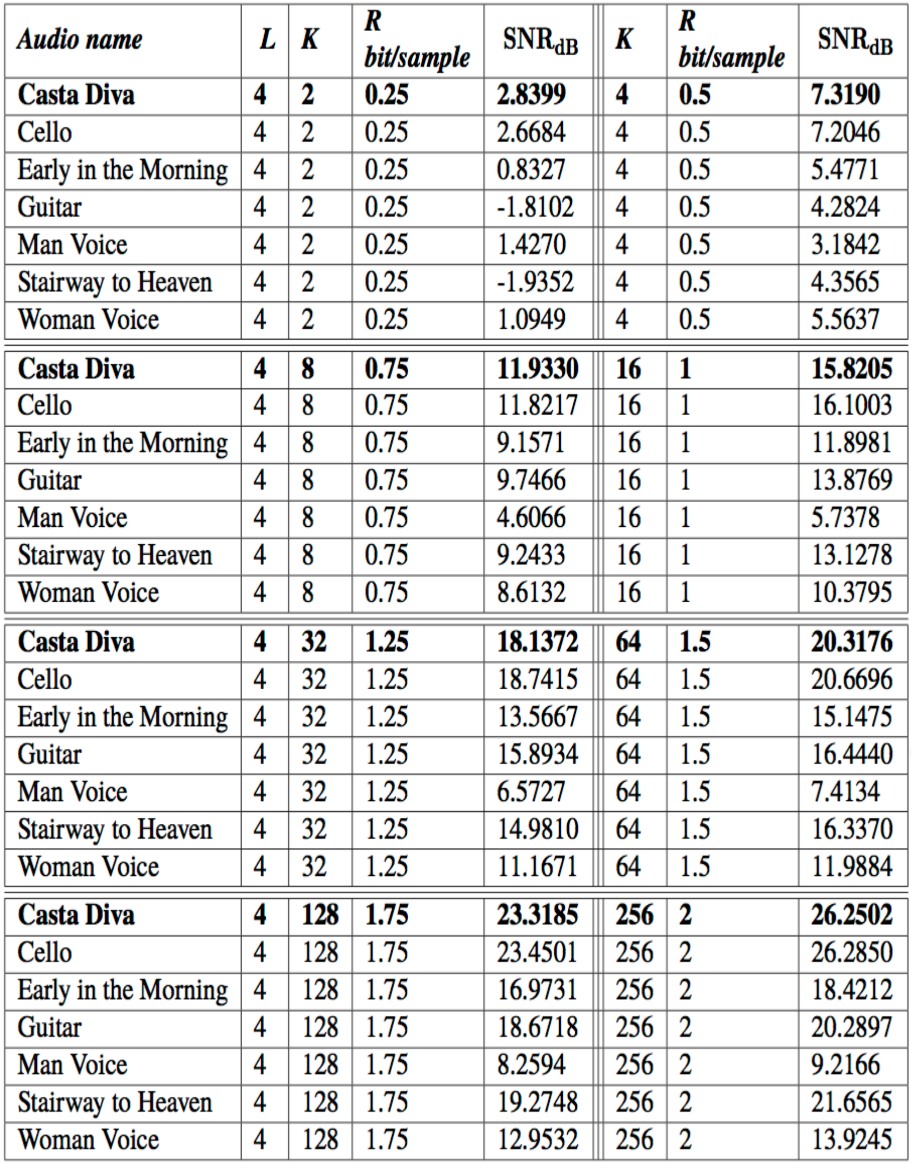
✓ Lower rate

- Lower *SNR*

Results: L=4, mono, training=*Casta Diva*

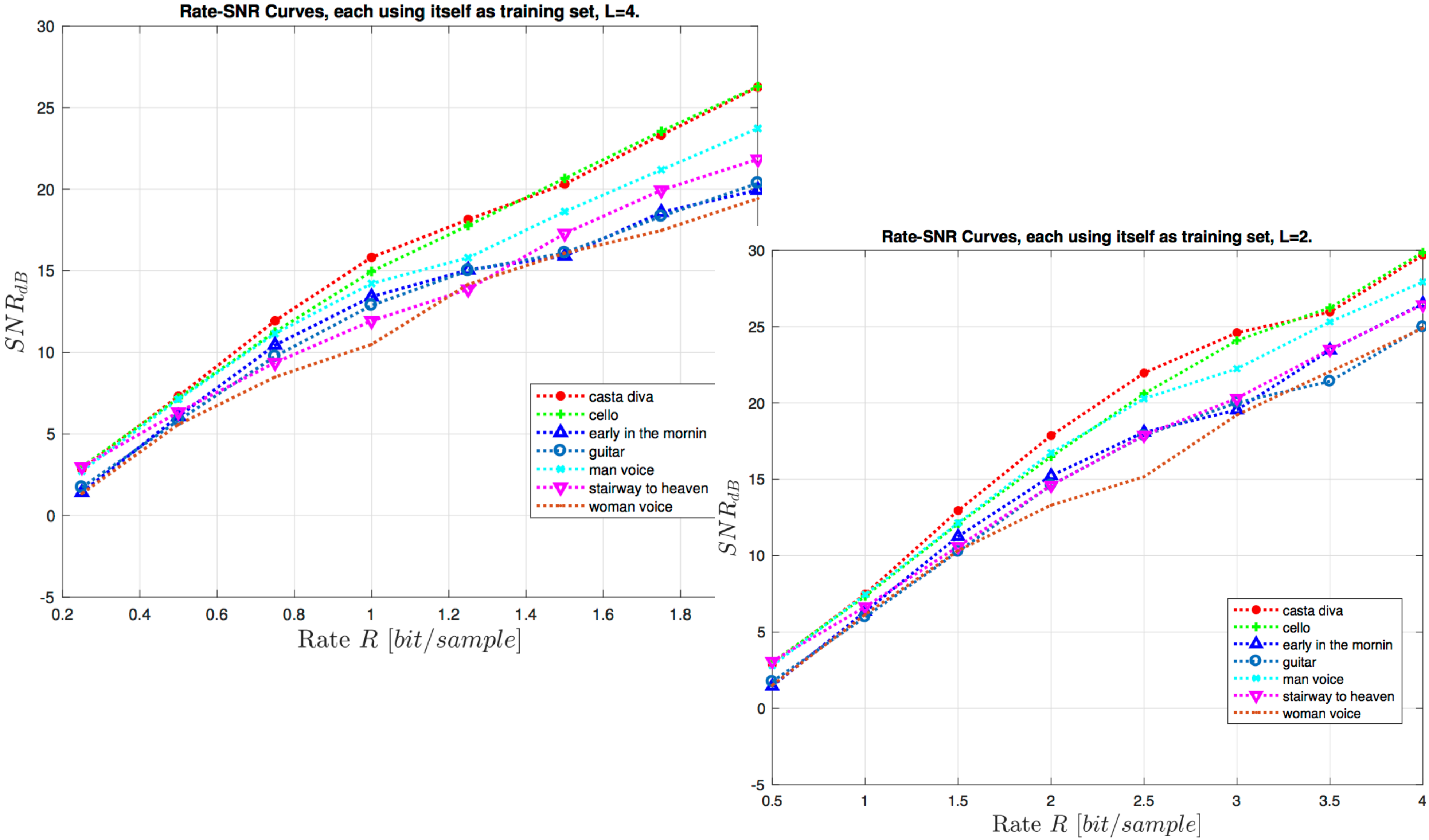


Audio name	*L*	*K*	*R* *bit/sample*	$\mathrm{SNR_{dB}}$	*K*	*R* *bit/sample*	$\mathrm{SNR_{dB}}$
Casta Diva	**2**	**2**	**0.5**	**2.8725**	**4**	**1**	**7.4876**
Cello	2	2	0.5	2.6925	4	1	7.3330
Early in the Morning	2	2	0.5	0.8704	4	1	5.6790
Guitar	2	2	0.5	-1.8154	4	1	4.3871
Man Voice	2	2	0.5	1.4508	4	1	3.2558
Stairway to Heaven	2	2	0.5	-1.9086	4	1	4.5466
Woman Voice	2	2	0.5	1.3290	4	1	6.1607
Casta Diva	**2**	**8**	**1.5**	**12.9501**	**16**	**2**	**17.8592**
Cello	2	8	1.5	12.5083	16	2	17.6690
Early in the Morning	2	8	1.5	10.0789	16	2	13.4206
Guitar	2	8	1.5	10.2144	16	2	15.5869
Man Voice	2	8	1.5	4.9777	16	2	6.1117
Stairway to Heaven	2	8	1.5	10.0850	16	2	15.3314
Woman Voice	2	8	1.5	10.3523	16	2	13.1654
Casta Diva	**2**	**32**	**2.5**	**21.9642**	**64**	**3**	**24.6008**
Cello	2	32	2.5	22.0293	64	3	24.9176
Early in the Morning	2	32	2.5	15.7615	64	3	17.2942
Guitar	2	32	2.5	19.3825	64	3	21.2062
Man Voice	2	32	2.5	6.9639	64	3	7.6450
Stairway to Heaven	2	32	2.5	19.2954	64	3	21.2993
Woman Voice	2	32	2.5	14.6991	64	3	15.4456
Casta Diva	**2**	**128**	**3.5**	**25.9558**	**256**	**4**	**29.6845**
Cello	2	128	3.5	26.5651	256	4	29.8972
Early in the Morning	2	128	3.5	18.6354	256	4	21.1628
Guitar	2	128	3.5	21.8688	256	4	23.9498
Man Voice	2	128	3.5	8.4725	256	4	9.4948
Stairway to Heaven	2	128	3.5	22.0431	256	4	25.8277
Woman Voice	2	128	3.5	15.9338	256	4	17.2472



Audio name	*L*	*K*	*R* *bit/sample*	$\mathrm{SNR_{dB}}$	*K*	*R* *bit/sample*	$\mathrm{SNR_{dB}}$
Casta Diva	**4**	**2**	**0.25**	**2.8399**	**4**	**0.5**	**7.3190**
Cello	4	2	0.25	2.6684	4	0.5	7.2046
Early in the Morning	4	2	0.25	0.8327	4	0.5	5.4771
Guitar	4	2	0.25	-1.8102	4	0.5	4.2824
Man Voice	4	2	0.25	1.4270	4	0.5	3.1842
Stairway to Heaven	4	2	0.25	-1.9352	4	0.5	4.3565
Woman Voice	4	2	0.25	1.0949	4	0.5	5.5637
Casta Diva	**4**	**8**	**0.75**	**11.9330**	**16**	**1**	**15.8205**
Cello	4	8	0.75	11.8217	16	1	16.1003
Early in the Morning	4	8	0.75	9.1571	16	1	11.8981
Guitar	4	8	0.75	9.7466	16	1	13.8769
Man Voice	4	8	0.75	4.6066	16	1	5.7378
Stairway to Heaven	4	8	0.75	9.2433	16	1	13.1278
Woman Voice	4	8	0.75	8.6132	16	1	10.3795
Casta Diva	**4**	**32**	**1.25**	**18.1372**	**64**	**1.5**	**20.3176**
Cello	4	32	1.25	18.7415	64	1.5	20.6696
Early in the Morning	4	32	1.25	13.5667	64	1.5	15.1475
Guitar	4	32	1.25	15.8934	64	1.5	16.4440
Man Voice	4	32	1.25	6.5727	64	1.5	7.4134
Stairway to Heaven	4	32	1.25	14.9810	64	1.5	16.3370
Woman Voice	4	32	1.25	11.1671	64	1.5	11.9884
Casta Diva	**4**	**128**	**1.75**	**23.3185**	**256**	**2**	**26.2502**
Cello	4	128	1.75	23.4501	256	2	26.2850
Early in the Morning	4	128	1.75	16.9731	256	2	18.4212
Guitar	4	128	1.75	18.6718	256	2	20.2897
Man Voice	4	128	1.75	8.2594	256	2	9.2166
Stairway to Heaven	4	128	1.75	19.2748	256	2	21.6565
Woman Voice	4	128	1.75	12.9532	256	2	13.9245

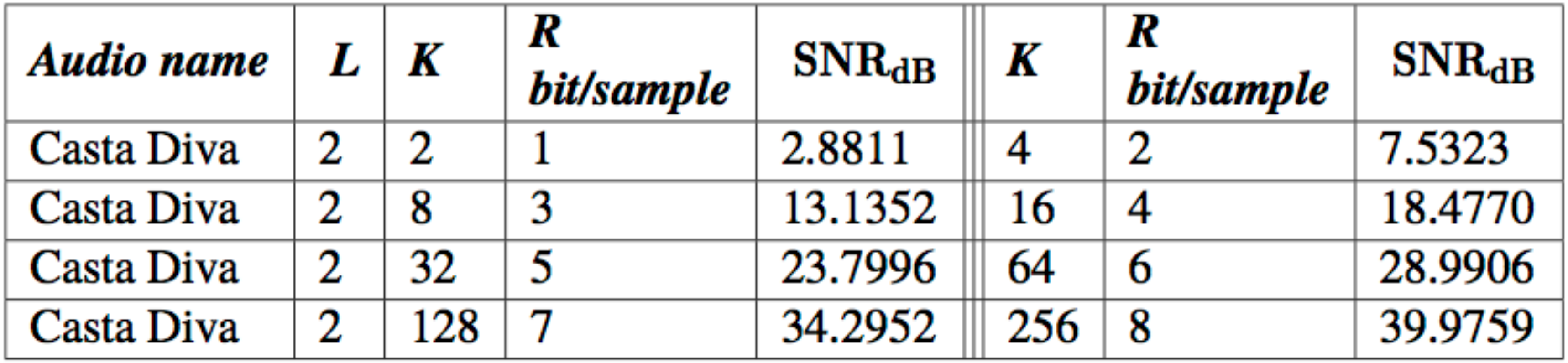
Results: L=4, mono, training=*itself*



Rate-SNR Curves, each using itself as training set, L=4.

Rate-SNR Curves, each using itself as training set, L=2.

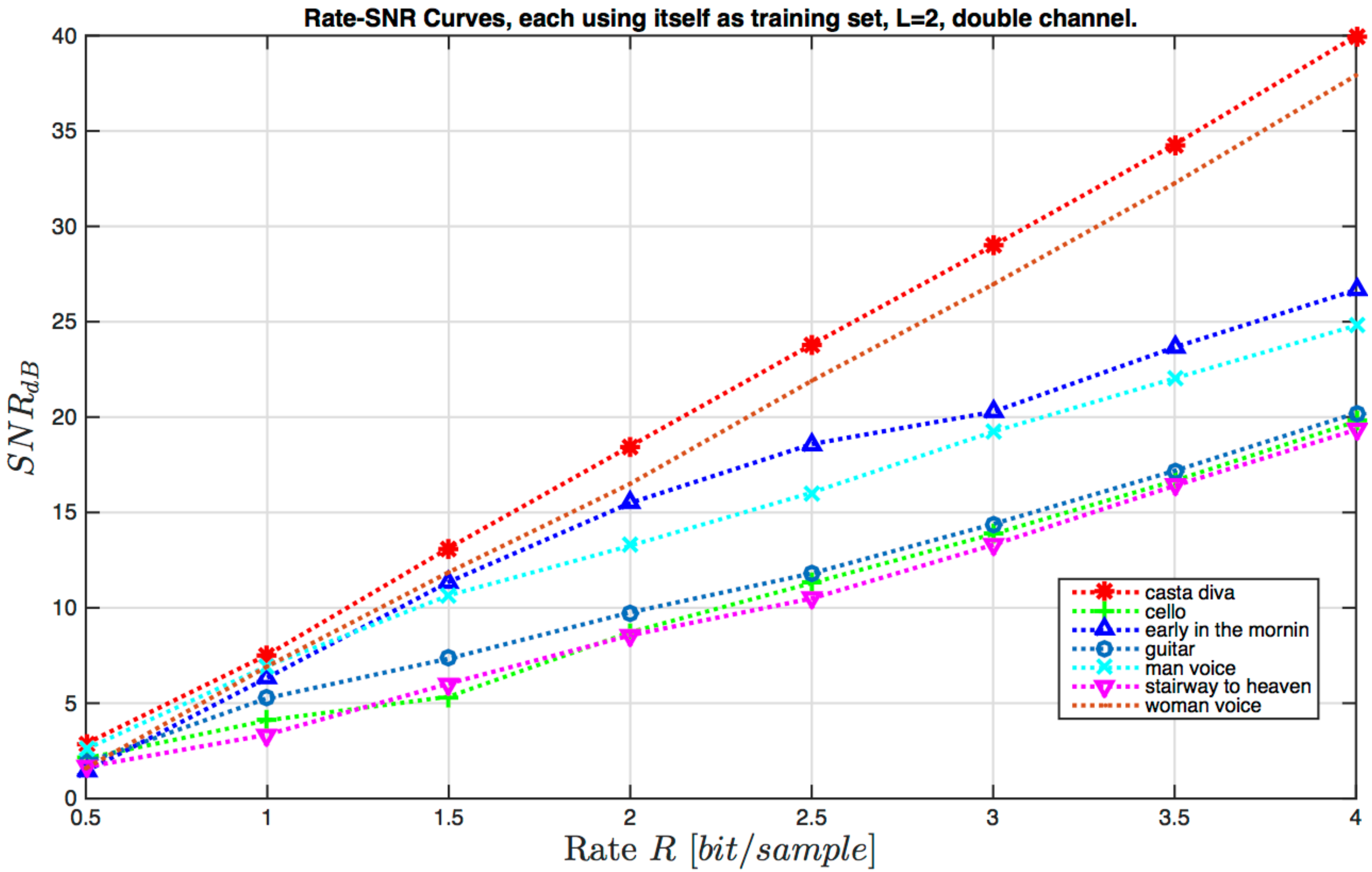
Results: L=2, double, training=*Casta Diva*



Audio name	*L*	*K*	*R* *bit/sample*	SNR$_{dB}$	*K*	*R* *bit/sample*	SNR$_{dB}$
Casta Diva	2	2	1	2.8811	4	2	7.5323
Casta Diva	2	8	3	13.1352	16	4	18.4770
Casta Diva	2	32	5	23.7996	64	6	28.9906
Casta Diva	2	128	7	34.2952	256	8	39.9759

- Trying to exploit the correlation between the two channels
- Values higher than before
- Rates are doubled because 2 channels

Results: L=2, double, training=*itself*



Some signals not improved because correlation between the channels is low

Summary

- Intro to coding techniques
- Vector Quantization
- LBG algorithm
- LBG applications:
  - L=2, mono, training with one audio or with the audio itself or with mixed audios
  - L=2, double, training with one audio or with the audio itself
  - L=4, mono, training with one audio or with the audio itself