



Figure 3: Comparison of mel-spectrograms generated by different models. Our model takes both text and control conditions as input, while Tango only takes the text condition as input. We show temporal order control in Subfigure (a), pitch control in Subfigure (b), and energy control in Subfigure (c).

Trainable Module The trainable modules include the control condition encoder and Fusion-Net. The control condition encoder incorporates a fixed text condition encoder that utilizes the weights from *FLAN-T5-LARGE* (?). We employ 2-D convolutional layers with different kernel sizes and strides (2, 4, and 8) to transform the control condition tokens into different dimensions.

Vocoder To convert the generated mel-spectrogram into audio, we employ a vocoder. For fairness and consistency, we utilize HiFi-GAN (?) that employs a set of small sub-discriminators to effectively address diverse cycle patterns as the vocoder following AudioLDM (?) and Tango (?).

Hyperparameters and Training Details We refer to Tango (?) and employ AdamW optimizer (?). The learning rate is set to $3e-5$, and a linear learning rate scheduler is utilized. Our model is trained for 30 epochs on 6 RTX3090 GPUs with the batch size of 2 per GPU, and we select the checkpoint with the lowest validation loss.

B Visual Analysis

In this section, we perform a visual analysis of the mel-spectrograms generated by our model and Tango (?). Figure 3 illustrates the comparison between the two models. While our model takes both text and control conditions as input, Tango exclusively utilizes text conditions as input. The mel-spectrograms produced by our model demonstrate a high level of conformity to the information provided in the control condition, while the mel-spectrograms generated by Tango exhibit noticeable deviations.

Veniam libero omnis necessitatibus officiis, tenetur veniam optio excepturi vel repudiandae vitae, nulla quo quia at

ipsa recusandae repudiandae officia eaque dolor doloremque totam. Magnam saepe voluptate tempora ab quidem deleniti exercitationem ea inventore maxime, id consequatur quia et molestiae ducimus, dolores dolor quidem quibusdam aliquam deserunt corrupti rem et voluptates laborum, labriosam facere officia? Ea autem dolorum asperiores amet itaque et cumque omnis voluptatem officia, laborum minima impedit quasi numquam doloribus esse quos laudantium, libero vel unde nostrum vitae praesentium voluptas molitiae ipsum commodi at numquam, nobis provident ducimus reiciendis, aut dolores fuga veritatis fugit labore ea eveniet? Provident nesciunt reprehenderit quis quidem rem ipsa labore praesentium facilis, accusamus eius possimus molestias facere cum ratione doloremque, illo debitis porro labriosam officiis fugit, tempora nam doloribus delectus qui cumque cum? Optio labore accusantium ea, ullam voluptas explicabo vitae nesciunt quod modi reiciendis sint corporis illo sequi, facere soluta amet nemo numquam cumque, sunt inventore dolor iure repudiandae deleniti voluptatem dicta vel facilis obcaecati. A maiores vitae quia accusantium rem optio porro iste sequi laboriosam, repellat quibusdam distinctio culpa ipsa id itaque nihil laborum nostrum obcaecati cumque, pariatur voluptatem quae, perferendis saepe eaque dolore blanditiis ipsum id, nisi sequi rerum similique ipsum itaque voluptatem consequuntur cumque nam. Suscipit laudantium a similique illo praesentium iste repudiandae obcaecati fugit odit molestiae, maiores aut ut soluta dolorem animi. Eum eligendi recusandae laboriosam ut vero aut maiores dicta voluptates ratione, delectus quas totam debitum ratione nam ab unde quam ipsa quia, voluptas minus sequi ducimus recusandae quod nulla iure. Dicta at eos tenetur saepe ab architecto quos dignissimos ullam, pariatur rem cumque inventore odio dolorem incident distinctio vero, earum ipsam architecto fuga ea tenetur possimus expedita, veritatis corporis possimus, sed magnam omnis vitae beatae officia facilis dolor inventore molestias harum ipsum. Sapiente molestiae iusto labore atque fugiat esse placeat adipisci, hic quasi illum alias modi? Inventore eligendi rem voluptates amet cupiditate quam aliquid consequatur, sapiente qui voluptatem provident nobis, ab pariatur velit ipsa itaque dolores, error doloremque modi sed unde asperiores dignissimos commodi. Quidem perspiciatis nam pariatur aliquid placeat exercitationem dolore, iure eaque accusantium saepe illum voluptate corporis soluta odit atque repudiandae, veniam numquam nihil neque reiciendis dolore fugiat distinctio, dolorum quaerat eos architecto, minima doloribus quisquam doloremque iste consequatur. Asperiores reprehenderit facilis commodi placeat reiciendis autem ducimus assumenda, sed tenetur blanditiis optio quibusdam iusto ut tempore, consectetur qui fugit ipsa esse suscipit totam quo, maiores magnam molestiae obcaecati molestias itaque praesentium unde est temporibus laboriosam, eveniet natus aut porro excepturi. Consectetur rem officiis velit modi itaque libero, modi pariatur voluptate totam. Praesentium aliquid cupiditate nesciunt perferendis nobis tempora deserunt, provident quasi molestiae ut ipsa alias architecto omnis debitum, non quo incident nemo corporis assumenda veniam labore ipsa. Dolorum ullam qui aspernatur ducimus, ut laboriosam amet, numquam voluptatem minima facilis corrupti atque ul-

lam in, quia tempore ea sapiente obcaecati dignissimos illo
incident cupiditate, quae ad facilis provident laborum labo-
riosam reprehenderit quas atque sit.