⚡️🚀המאמר היומי של מייק 07.08.24: ⚡️🚀

Language Model Can Listen While Speaking

המאמר שמשך את תשומת ליבי בגלל שמו הקליט. המאמר מציע ארכיטקטורה של מודל Speech Language Model או SLM שיודע להקשיב תוך כדי שהוא מדבר, כלומר מודל full duplex (מושג מתחום התקשורת). בדרך כלל ל- SLM יש שני משטר עבודה: הקשבה או דיבור, כלומר המודל או מדבר או מקשיב. המאמר מעשיר את מרחב היכולות של SLM ומצייד אותו ביכולת להקשיב תוך כדי שהוא מדבר. מעניין שהמודל גם יכול לעצור אם הוא מזהה שיש דיבור (לא רעש) ומגיב עליו (בדיבור) לאחר מכן.

הארכיטקטורה של המודל המוצע LSLM מורכב מרכיבים סטנדרטיים. יש מודל שקולט אות דיבור, מחלק אותו לטוקנים (האות במקטעי זמן שונים) מקודד אותו לוקטור אמבדינג ומאזין אותו לדקודר. תפקיד הדקודר הוא לקחת בחשבון את ייצוג של טוקני הדיבור שנקלטו קודם וגם ייצוג טוקני הדיבור שנוצרו על ידי המודל כדי ליצור את הפלט הבא (אות הדיבור) של המודל. ֿכאמור לפעמים הדקודר מחליט שהוא צריך לעבור למצב האזנה ולפעמים הוא צריך לעבור למצב הדיבור.

כלומר הדקודר במקרה הזה הוא vocoder המקבל כקלט את אות הדיבור הנקלט בנוסף לאות הדיבור המגונרט על ה-vocoder עצמו לפני.

https://arxiv.org/pdf/2408.02622